A Unifying Variational Perspective on Some Fundamental Information Theoretic Inequalities

Sangwoo Park, Erchin Serpedin, and Khalid Qaraqe

Abstract

This paper proposes a unifying variational approach for proving and extending some fundamental information theoretic inequalities. Fundamental information theory results such as maximization of differential entropy, minimization of Fisher information (Cramér-Rao inequality), worst additive noise lemma, entropy power inequality (EPI), and extremal entropy inequality (EEI) are interpreted as functional problems and proved within the framework of calculus of variations. Several applications and possible extensions of the proposed results are briefly mentioned.

Index Terms

Maximizing Entropy, Minimizing Fisher Information, Worst Additive Noise, Entropy Power Inequality, Extremal Inequality, Calculus of Variations

I. Introduction

In the information theory realm, it is well-known that given the second-order moment (or variance), a Gaussian density function maximizes the differential entropy. Similarly, given the second-order moment, the Gaussian density function minimizes the Fisher information, a result which is referred to as the Cramér-Rao inequality in the signal processing literature. Surprisingly, the proofs proposed in the literature for these fundamental results are relatively quite diverse, and no unifying feature exists. Since differential entropy or Fisher information is a functional with respect to a probability density function, the most natural way to establish these results is by approaching them from the perspective of functional analysis. Although some of these results have been dealt with partially or not all within the framework of calculus of variations, this paper presents a unifying variational framework to address these results as well as

Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX 77843-3128, USA, e-mail: serpedin@ece.tamu.edu

numerous other fundamental information theoretic results. A number of challenging information theoretic inequalities such as the entropy power inequality (EPI) [1] and the extremal entropy inequality (EEI) [2] can be dealt with successfully in the proposed framework of functionals. Furthermore, the proposed variational calculus perspective presents usefulness in establishing other novel results and new extensions for the existing information theoretic inequalities.

The main theme of this paper is to illustrate how some tools from calculus of variations can be used successfully to prove some of the fundamental information theoretic inequalities, which have been widely used in information theory and other fields. The proposed variational approach provides alternative proofs for some of the fundamental information theoretic inequalities and enables finding novel extensions of the existing results. However, more importantly is the fact that the proposed variational approach offers a potential guideline for finding the optimal solution for many other open problems. In addition, as a general feature, the proposed variational approach enables finding simpler solutions to some quite challenging results such as EPI or EEI.

Variational calculus techniques have been used with great success in solving important problems in image processing and computer vision [4] such as image reconstruction (denoising, deblurring), inverse problems, and image segmentation. Recently, variational techniques were also advocated by Scutari and Palomar for optimization of multiuser communication systems [5]-[7], for deriving analytical wireless channel models using the maximum entropy principle when only limited information about the environment is available [8]-[10] and for designing optimal training sequences for radar and sonar applications [11]-[13]. Maximum entropy principle found also applications in spectral estimation (e.g., Burg's maximum entropy spectral density estimator [1]) and Bayesian statistics [14].

The major results of this paper are enumerated as follows. First, using calculus of variations, the maximizing differential entropy and minimizing Fisher information theorems are proved under the classical (standard) assumptions found in the literature as well as under a different set of assumptions. It is shown that a Gaussian density function maximizes the differential entropy but it minimizes the Fisher information, given the second-order moment. It is also shown that a half normal density function maximizes the differential entropy over the set of non-negative random variables, given the second-order moment. Furthermore, it is shown that a half normal density function minimizes the Fisher information over the set of non-negative random variables, provided that the regularity condition is ignored and the second-order moment is given. It is also shown that a chi density function minimizes the Fisher information over the set of non-negative random variables, under the assumption that the regularity condition holds and the second-order moment is given.

Second, a novel proof of the worst additive noise lemma [20] is provided in the proposed functional framework. Previous proofs of the worst additive noise lemma were based on Jensen's inequality or data processing inequality [20], [21]. Unlike the previous proofs, our approach is purely based on calculus of variations techniques, and both the scalar and vector versions of the lemma are treated.

Third, EPI is proved based on calculus of variations. We first re-cast EPI into a functional problem. Then, the necessary optimal solutions for the functional problem are found using Euler's equation and exploiting the necessary conditions for the existence of the optimal solution for the considered functional problem. In a scalar version of EPI, the necessarily optimal solution, which is the Gaussian density function, is actually sufficient since only the Gaussian density function satisfies the Euler's equation. This is one of the main benefits using calculus of variations since it allows finding global optimal solutions simply by checking the set of solutions imposed by the set of necessary conditions. In a vector version of EPI, Euler's equation only shows that the Gaussian density functions are necessarily optimal, since the covariance matrices of the optimal solutions are not determined. However, this information alone–i.e., the fact that the optimal solutions are Gaussian–is enough to prove EPI.

Finally, EEI is studied from the perspective of a functional problem. The main advantage of the proposed new proof is that neither the channel enhancement technique and EPI, used in [2], nor the equality condition of data processing inequality and the technique based on the moment generating functions, adopted in [27], are required. Using the unified argument based on calculus of variations, EEI is simply proved herein paper.

The rest of this paper is organized as follows. Some variational calculus preliminary results and their corollaries are first reviewed in Section II. Maximizing differential entropy theorem and minimizing Fisher information theorem (Cramér-Rao inequality) are proved in Section III. In Section IV, the worst additive noise lemma is introduced and proved based on calculus of variations. EPI and EEI are proved in Sections V and VI, respectively. In Section VII, some applications of the addressed information theoretic inequalities are briefly mentioned. Finally, Section VIII concludes this paper.

II. Some Preliminary Calculus of Variations Results

In this section, we will review some of the fundamental results from variational calculus, and establish the concepts, notations and results that will be used constantly throughout the rest of the paper. These results are standard and therefore will be described briefly without further details. The readers might consult any book on calculus of variations such as [15], [16], [17].

Definition 1: A functional $U[f_x]$ is defined as

$$U[f_x] = \int_a^b K(x, f_x, f_x')dx, \tag{1}$$

which is defined on the set of continuous functions. The function f_x is assumed to have continuous first-order derivative in $[a, b]$ and to satisfy the boundary conditions $f_x(a) = A_x$ and $f_x(b) = B_x$. The functional $K(\cdot, \cdot, \cdot)$ is also assumed to have continuous first-order and second-order (partial) derivatives with respect to (wrt) all of its arguments. Also, notation f_x' denotes the first-order derivative wrt x.

Definition 2: The increment of a functional $U[f_x]$ is defined as

$$\Delta U[h_x] = U[f_x + h_x] - U[f_x], \tag{2}$$

where the function h_x is the increment, and it is independent of the function f_x.

Definition 3: Suppose that, given f_x,

$$\Delta U\left[h_x\right] = \varphi\left[h_x\right] + \epsilon \|h_x\|, \tag{3}$$

where $\varphi\left[h_x\right]$ is a linear functional, ϵ goes to zero as $\|h_x\|$ approaches zero, and $\|\cdot\|$ denotes a norm and it is defined as

$$\|f_x\| = \sum_{i=0}^{n} \max_{a \le x \le b} \left| f_x^{(i)}(x) \right|, \tag{4}$$

where $f_x^{(i)}(x) = (d^i/dx^i) f_x(x)$, and summation upper index n varies depending on the normed linear space considered (e.g., if the normed linear space consists of all continuous functions $f_x(x)$– which have continuous first-order derivative–defined on an interval $[a, b]$, $\|f_x\| = \max_{a \le x \le b} |f_x(x)| + \max_{a \le x \le b} |f_x'(x)|$, and in this case $n = 1$). Then, the functional $U\left[f_x\right]$ is said to be differentiable, and the major part of the increment $\varphi\left[h_x\right]$ is called the (first-order) variation of the functional $U\left[f_x\right]$ and it is expressed as $\delta U\left[f_x\right]$.

Based on Definitions 1, 2, 3 and Taylor's theorem (see [15]), the first-order and the second-order variations of a functional $U\left[f_x\right]$ are expressed as

$$\delta U\left[f_x\right] = \int \left[K_{f_x}'\left(x, f_x, f_x'\right) h_x(x) + K_{f_x'}'\left(x, f_x, f_x'\right) h_x'(x)\right] dx, \tag{5}$$

$$\begin{aligned}
\delta^2 U\left[f_x\right] &= \frac{1}{2} \int \Big[K_{f_x f_x}''\left(x, f_x, f_x'\right) h_x(x)^2 + 2K_{f_x f_x'}''\left(x, f_x, f_x'\right) h_x(x) h_x'(x) \\
&\quad + K_{f_x' f_x'}''\left(x, f_x, f_x'\right) h_x'(x)^2\Big] dx \\
&= \frac{1}{2} \int \left[K_{f_x' f_x'}'' {h_x'}^2 + \left(K_{f_x f_x}'' - \frac{d}{dx} K_{f_x f_x'}''\right) {h_x}^2\right] dx,
\end{aligned} \tag{6}$$

where K'_{f_X} and $K'_{f'_X}$ are the first-order partial derivatives wrt f_X and f'_X, respectively, $K''_{f_X f'_X}$ is the second-order partial derivative wrt f_X and f'_X, $K''_{f_X f_X}$ is the second-order partial derivative wrt f_X, and $K''_{f'_X f'_X}$ is the second-order partial derivative wrt f'_X.[1]

Theorem 1 ([15]): A necessary condition for the functional $U[f_X]$ in (1) to have an extremum (or, local optimum) for a given function f_{X^*} is the following:

$$\delta U[f_{X^*}] = 0, \tag{7}$$

for all admissible h_X. This implies

$$K'_{f_{X^*}} - \frac{d}{dx} K'_{f'_{X^*}} = 0, \tag{8}$$

a result which is known as Euler's equation. When the functional in (1) includes multiple functions (e.g., $f_{X_1}, \ldots, f_{X_n}$) and multiple integrals wrt $x_1, \ldots, x_n$, then Euler's equation in (8) is changed to

$$K'_{f_{X_i^*}} - \sum_{j=1}^{n} \frac{d}{dx_j} K'_{f'_{X_i^*}} = 0, \qquad i = 1, \ldots, n. \tag{9}$$

In particular, when the functional does not depend on the first-order derivative of the functions $f_{X_1}, \ldots, f_{X_n}$, the equation in (9) is simplified to

$$K'_{f_{X_i^*}} = 0, \qquad i = 1, \ldots, n. \tag{10}$$

Proof: Details of the proof of this theorem can be found e.g., in [15]. ■

Theorem 2 ([15]): A necessary condition for the functional $U[f_X]$ in (1) to have a minimum for a given f_{X^*} is the following:

$$\delta^2 U[f_{X^*}] \geq 0, \tag{11}$$

for all admissible h_X. This implies

$$K''_{f'_{X^*} f'_{X^*}} \geq 0. \tag{12}$$

In particular, when the functional in (1) does not depend on the first-order derivative of the function f_X, the equation in (12) changes into

$$K''_{f_{X^*} f_{X^*}} \geq 0. \tag{13}$$

[1]Throughout the paper, the arguments of functionals or functions are omitted unless the arguments are ambiguous or confusing.

When the functional in (1) includes multiple functions (e.g., $f_{X_1}, \ldots, f_{X_n}$) and multiple integrals wrt $x_1, \ldots, x_n$, then the equation in (13) is changed into the positive semi-definiteness of the following matrix:

$$\begin{bmatrix} K''_{f_{X_1} f_{X_1}} & \cdots & K''_{f_{X_1} f_{X_n}} \\ \vdots & \ddots & \vdots \\ K''_{f_{X_n} f_{X_1}} & \cdots & K''_{f_{X_n} f_{X_n}} \end{bmatrix} \geq 0. \tag{14}$$

Proof: The inequality in (13) is easily derived from the inequality in (12) since $K''_{f'_X f'_X}$ and $K''_{f_X f'_X}$ are vanishing in (6) when the functional in (1) does not depend on the first-order derivative of the function f_X. Additional details of the proof can be found in [15]. ■

Theorem 3 ([15]): Given the functional

$$U[f_X, f_Y] = \int_a^b K(x, f_X, f_Y, f'_X, f'_Y) dx, \tag{15}$$

assume that the admissible functions satisfy the following conditions:

$$f_X(a) = A_X, \quad f_X(b) = B_X, \quad f_Y(a) = A_Y, \quad f_Y(b) = B_Y,$$

$$k(x, f_X, f_Y) = 0, \tag{16}$$

$$L[f_X, f_Y] = \int_a^b \tilde{L}(x, f_X, f_Y, f'_X, f'_Y) dx = l, \tag{17}$$

where a, b, A_X, B_X, A_Y, B_Y, and l are constants, and $U[f_X, f_Y]$ is assumed to have an extremum for $f_X = f_{X^*}$ and $f_Y = f_{Y^*}$.

If f_{X^*} and f_{Y^*} are not extremals of $L[f_X, f_Y]$, or $k'_{f_{X^*}}$ and $k'_{f_{Y^*}}$ do not vanish simultaneously at any point in (16), there exists a constant λ or a function $\lambda(x)$ such that f_{X^*} and f_{Y^*} are extremals of the functional

$$\int_a^b \Big(K(x, f_X, f_Y, f'_X, f'_Y) + \lambda \tilde{L}(x, f_X, f_Y, f'_X, f'_Y) + \lambda(x) k(x, f_X, f_Y) \Big) dx. \tag{18}$$

Based on Theorem 3, the following corollary is derived.

Corollary 1: Given the functional

$$U[f_X, f_Y] = \int_a^b \int_a^b K(x, y, f_X, f_Y) dx dy, \tag{19}$$

assume that the admissible functions satisfy the following conditions:

$$f_X(a,a) = A_X, \quad f_X(b,b) = B_X, \quad f_Y(a) = A_Y, \quad f_Y(b) = B_Y, \quad k(x, y, f_X, f_Y) = 0,$$

$$L[f_X, f_Y] = \int_a^b \int_a^b \tilde{L}(x, y, f_X, f_Y) dx dy = l, \tag{20}$$

where a, b, A_X, B_X, A_Y, and B_Y are constants, f_X is a function of both x and y, f_Y is a function of y. The functional $k(y, f_X, f_Y)$ is defined as $g(y, f_Y) - \int_a^b \tilde{k}(x, y, f_X)dx$, where $g(y, f_Y)$ is a functional of f_Y and $\tilde{k}(x, y, f_X)$ is a functional of f_X. And, $U[f_X, f_Y]$ is assumed to have an extremum for $f_X = f_{X^*}$ and $f_Y = f_{Y^*}$.

Unless f_{X^*} and f_{Y^*} are extremals of $L[f_X, f_Y]$, or k'_{f_X} and k'_{f_Y} simultaneously vanish at any point of $k(x, y, f_X, f_Y)$, there exists a constant λ or a function $\lambda(y)$ such that $f_X = f_{X^*}$ and $f_Y = f_{Y^*}$ is an extremal of the functional

$$\int_a^b \left\{ \left(\int_a^b \left[K(x, y, f_X, f_Y) + \lambda \tilde{L}(x, y, f_X, f_Y) - \lambda(y) k(x, y, f_X) \right] dx \right) + \lambda(y) g(y, f_Y) \right\} dy. \tag{21}$$

Proof: This corollary is a simple extension of Theorem 3 for multiple integrals. Therefore, the detailed proof is omitted. ■

Based on Theorems 1, 2 and Corollary 1, we can derive the following corollary, which will be mainly used throughout this paper.

Corollary 2: Based on the functional defined in (21), the following necessary conditions are derived for the optimal solutions f_{X^*} and f_{Y^*}:

$$K'_{f_{X^*}}(x, y, f_{X^*}, f_{Y^*}) - \lambda \tilde{L}'_{f_{X^*}}(x, y, f_{X^*}, f_{Y^*}) - \lambda(y) k'_{f_{X^*}}(x, y, f_{X^*}) = 0, \tag{22}$$

$$\int K'_{f_{Y^*}}(x, y, f_{X^*}, f_{Y^*}) - \lambda \tilde{L}'_{f_{Y^*}}(x, y, f_{X^*}, f_{Y^*}) dx + \lambda(y) g'_{f_{Y^*}}(y, f_{Y^*}) = 0, \tag{23}$$

and the matrix

$$\begin{bmatrix} G''_{f_{X^*}, f_{X^*}} & G''_{f_{X^*}, f_{Y^*}} \\ G''_{f_{Y^*}, f_{X^*}} & G''_{f_{Y^*}, f_{Y^*}} \end{bmatrix}, \tag{24}$$

where the functional G is defined as

$$G(x, y, f_{X^*}, f_{Y^*}) = K(x, y, f_{X^*}, f_{Y^*}) - \lambda \tilde{L}(x, y, f_{X^*}, f_{Y^*}) - \lambda(y) k(x, y, f_{X^*}) + \lambda(y) g(y, f_{Y^*}) q(x),$$

and $q(x)$ is a function which satisfies $\int_a^b q(x)dx = 1$, is positive definite.

Proof: The equations in (22) and (23) are derived from the first-order variation condition in Theorem 1. Namely, the equations in (22) and (23) are Euler's equations for multiple integrals. The positive definiteness of the matrix in (24) is derived from the second-order variation condition in Theorem 2. Namely, this is the same as the one in (14). Since the proof is straightforward, the details of the proof are omitted here. ■

III. MAX ENTROPY AND MIN FISHER INFORMATION

This simple but significant result–given the second-order moment (or variance) of a random variable, a Gaussian density function maximizes the differential entropy while it minimizes the Fisher information–is well-known. However, its complete rigorous proof can hardly be found. In this section, using calculus of variations, complete rigorous proofs will be provided.

Theorem 4 ([1]): Given (the first-order) and the second-order moments of a random variable X, differential entropy of the random variable X is maximized when X is Gaussian, i.e.,

$$h(X) \;\; \leq \;\; h(X_G), \tag{25}$$

where $h(\cdot)$ denotes differential entropy, and X_G is a Gaussian random variable whose (first-order) and second-order moments are identical to the one of X.

Proof: In [1], the proof relies on calculus of variations to find the first-order necessary condition, which confirms necessary optimal solutions. However, the first-order necessary condition shows neither whether the solutions are local minimal or local maximal nor whether the solutions are locally optimal or globally optimal. Therefore, an additional technique, the Kullback-Leibler divergence, was used to prove that the necessary solution globally maximizes the differential entropy. Unlike this proof, by confirming both the first-order and the second-order necessary conditions, one can show that the optimal solution is a local maximal. Then, it can be shown that the local maximal is an actual global maximum achieving solution by proving that the local maximal is the only solution in the feasible set. Therefore, one can prove Theorem 4 solely based on calculus of variations arguments. See Appendix A for the details of the proof.

Remark 1: Even though the proposed proof is performed assuming constraints on the first-order and the second-order moments, the constraint on the first-order moment is not necessary. This will be shown in the proof of Theorem 5, which is the vector version of this theorem.

■

Similar to Theorem 4, given a correlation matrix (or a covariance matrix), a multi-variate Gaussian density function maximizes the differential entropy as shown by the following theorem.

Theorem 5 ([1], [21]): Given (a mean vector $\boldsymbol{\mu}_X$) and a correlation matrix $\boldsymbol{\Omega}_X$, a Gaussian random vector maximizes the differential entropy, i.e.,

$$h(\mathbf{X}) \;\; \leq \;\; h(\mathbf{X}_G), \tag{26}$$

where $h(\cdot)$ denotes differential entropy, $\mathbf{X}$ is an arbitrary but fixed random vector with the correlation matrix $\boldsymbol{\Omega}_X$, and $\mathbf{X}_G$ is a Gaussian random vector whose correlation matrix is identical to the one of $\mathbf{X}$.

Proof: See Appendix B.

Remark 2: The proposed proof is different from the ones mentioned in [1], [21] in the sense that the proposed proof relies only on variational calculus techniques. Moreover, from the proposed proof, one can observe that the constraint related to the first-order moment is not necessary.

Remark 3: Depending on the existence of the constraint related to the mean vector, the mean of the optimal Gaussian density function will change. However, the constraint on the mean vector is not necessarily required. Details of the proof are presented in Appendix B.

■

If we only consider non-negative random variables, a Gaussian random variable is not the solution which maximizes the differential entropy. The following theorem shows that a half-normal random variable maximizes the differential entropy over the set of non-negative random variables.

Theorem 6: Given an arbitrary but fixed non-negative random variable X and a half-normal random variable $X_{_{HN}}$, whose second moments are identical to those of X, then the following relationship holds:

$$h(X) \;\; \leq \;\; h(X_{_{HN}}), \tag{27}$$

where $h(\cdot)$ denotes differential entropy.

Proof: See Appendix C. ■

Similar to Theorems 4, 5, and 6, we can find a probability density function, which minimizes the Fisher information.

Theorem 7 (Cramér-Rao Inequality): Given (the first-order moment $\mu_{_X}$) and the second-order moment $m^2_{_X}$, a Gaussian random variable $X_{_G}$ minimizes Fisher information, i.e.,

$$J(X) \;\; \geq \;\; J(X_{_G}), \tag{28}$$

where X is an arbitrary but fixed random variable with the first-order moment $\mu_{_X}$ and the second-order moment $m^2_{_X}$. Notation $J(\cdot)$ denotes the Fisher information, and it is defined as

$$J(X) \;\; = \;\; \int \left(\frac{\frac{d}{dx} f_X(x)}{f_X(x)} \right)^2 f_X(x) dx.$$

Proof: See Appendix D.

Remark 4: Even though several proofs of this theorem have been proposed in the literature, this is the first rigorous proof of this theorem based on calculus of variations techniques.

■

Theorem 7 can be generalized to random vectors as shown in the following theorem.

Theorem 8 (Cramér-Rao Inequality (a vector version)): Given an arbitrary but fixed random vector $\mathbf{X}$ and a Gaussian random vector $\mathbf{X}_G$, whose mean vectors and correlation matrices are identical, respectively,

$$\mathbb{J}(\mathbf{X}) \quad \succeq \quad \mathbb{J}(\mathbf{X}_G), \tag{29}$$

where $\mathbb{J}(\cdot)$ denotes Fisher information matrix, and it is defined as

$$\mathbb{J}(\mathbf{X}) = \begin{bmatrix} s_{11} & \cdots & s_{1n} \\ \vdots & \ddots & \vdots \\ s_{n1} & \cdots & s_{nn} \end{bmatrix}, \tag{30}$$

$$s_{ij} = \int \left(\frac{\frac{d}{dx_i} f_X(\mathbf{x})}{f_X(\mathbf{x})} \right) \left(\frac{\frac{d}{dx_j} f_X(\mathbf{x})}{f_X(\mathbf{x})} \right) f_X(\mathbf{x}) d\mathbf{x}.$$

Proof: See Appendix E. ■

Similar to Theorem 7, a half-normal and a chi density function minimize the Fisher information over the set of non-negative random variables as shown in the following two theorems.

Theorem 9: Assume that the regularity condition for Fisher information is ignored. Given an arbitrary but fixed non-negative random variable X and a half-normal random variable X_{HN}, whose second order moments are identical to those of X, then the following inequality holds:

$$J(X) \quad \geq \quad J(X_{HN}), \tag{31}$$

where $J(\cdot)$ denotes Fisher information. The regularity condition is the following relationship:

$$\int \frac{d}{dx} f(x) dx = 0. \tag{32}$$

Proof: See Appendix F. ■

Theorem 10 ([18]): Assume next that random variables, which satisfy the regularity condition in (32), are considered. Given an arbitrary but fixed non-negative random variable X and a chi-distributed random variable X_C, whose second-order moments are identical to those of X, then the following inequality holds:

$$J(X) \quad \geq \quad J(X_C), \tag{33}$$

where $J(\cdot)$ stands for the Fisher information.

Proof: Unlike the proof in [18], by considering the first-order and the second-order moments instead of variance, we obtain the convex constraint sets. Since Fisher information is a strictly convex functional with respect to a probability density function, the variational problem is convex, and hence has an unique solution. The details of the proof are deferred to Appendix G. ■

IV. Worst Additive Noise Lemma

Worst additive noise lemma was introduced and exploited in several references [20], [21], [23], and it has been widely used in numerous applications. One of the main applications of the worst additive noise lemma pertains to the calculation of channel capacity under several different wireless communications scenarios such as the Gaussian MIMO broadcasting channel, Gaussian MIMO wire-tap channel, etc. In this section, the worst additive noise lemma for both random variables and random vectors will be proved solely based on calculus of variations arguments.

Theorem 11: Assume X is an arbitrary but fixed random variable and X_G is a Gaussian random variable, whose second-order moment is identical to that of X, which it is denoted as m_X^2. Given a Gaussian random variable W_G, which is independent of both X and X_G, with the second-order moment m_W^2, then the following relationship holds:

$$I(X + W_G; W_G) \geq I(X_G + W_G; W_G), \tag{34}$$

where $I(\cdot;\cdot)$ denotes mutual information.

Proof: The details of the proof are deferred to Appendix H. ■

Similarly, Theorem 11 can be generalized to random vectors as shown in the following theorem.

Theorem 12: Assume $\mathbf{X}$ is an arbitrary but fixed random vector and $\mathbf{X}_G$ is a Gaussian random vector, whose correlation matrix is identical to that of $\mathbf{X}$, and it is denoted as $\mathbf{\Omega}_X$. Given a Gaussian random vector $\mathbf{W}_G$, which is independent of both $\mathbf{X}$ and $\mathbf{X}_G$, with the correlation matrix $\mathbf{\Omega}_W$, then the following relation holds:

$$I(\mathbf{X} + \mathbf{W}_G; \mathbf{W}_G) \geq I(\mathbf{X}_G + \mathbf{W}_G; \mathbf{W}_G). \tag{35}$$

Proof: Our novel proof is entirely based on calculus of variations arguments. The summary of our proof is the following. First, we construct a variational problem, which represents the inequality in (35) and required constraints in a functional form. Second, using the first-order variation condition, we find necessary optimal solutions, which satisfy Euler's equation. Third, using the second-order variation condition, we show that the optimal solutions are necessarily local minima. Finally, we prove that the local minimum is also global. The details of the proof are presented to Appendix I. ■

V. Entropy Power Inequality

Entropy power inequality (EPI) is a powerful result that found applicability in determining the capacity of scalar Gaussian broadcast channel [24], the capacity of Gaussian MIMO broadcast channel [2], [19],

the secrecy capacity of Gaussian wire-tap channel [25], [27], etc., in conjunction with Fano's inequality and additional techniques such as the ones proposed in [19], [27]. In this section, we will prove several versions of EPI using calculus of variations techniques.

Theorem 13 (Entropy Power Inequality): For two independent random variables X and W, whose entropies and second-order moments are finite,

$$h(a_X X + a_W W) \geq a_X^2 h(X) + a_W^2 h(W), \tag{36}$$

where $a_X^2 + a_W^2 = 1$. The equality holds if and only if X and W are Gaussian random variables.

Proof: See Appendix J. ■

Theorem 14 (Entropy Power Inequality): For two independent random vectors $\mathbf{X}$ and $\mathbf{W}$, with finite entropies and correlation matrices, the following relation holds:

$$h(a_X \mathbf{X} + a_W \mathbf{W}) \geq a_X^2 h(\mathbf{X}) + a_W^2 h(\mathbf{W}), \tag{37}$$

where $a_X^2 + a_W^2 = 1$. The equality holds if and only if $\mathbf{X}$ and $\mathbf{W}$ are Gaussian random vectors and their covariance matrices $\mathbf{\Sigma}_X$ and $\mathbf{\Sigma}_W$ are identical.

Proof: See Appendix K. ■

VI. Extremal Entropy Inequality

Extremal entropy inequality, motivated by multi-terminal information theoretic problems such as the vector Gaussian broadcast channel and the distributed source coding with a single quadratic distortion constraint, was proposed by Liu and Viswanath [2]. It is an entropy power inequality which includes a covariance constraint. Because of the covariance constraint, the extremal entropy inequality could not be proved directly by using the classical EPI. Therefore, new techniques ([19], [27]) were adopted in the proofs reported in [2], [27]. In this section, the extremal entropy inequality will be proved using calculus of variations.

Theorem 15: Assume that μ is an arbitrary but fixed constant, where $\mu \geq 1$, and r^2 is a positive constant. A Gaussian random variable W_G with variance σ_W^2 is assumed to be independent of an arbitrary random variable X, with variance $\sigma_X^2 \leq r^2$. Then, there exists a Gaussian random variable X_G^* with variance $\sigma_{X^*}^2$ which satisfies the following inequality:

$$h(X) - \mu h(X + W_G) \leq h(X_G^*) - \mu h(X_G^* + W_G), \tag{38}$$

where $\sigma_{X^*}^2 \leq r^2$.

Proof: See Appendix L. ■

Theorem 15 can be generalized for random vectors as shown in the following two theorems.

Theorem 16: Assume that μ is an arbitrary but fixed constant, where $\mu \geq 1$, and $\mathbf{\Sigma}$ is a positive semi-definite matrix. A Gaussian random vector $\mathbf{W}_{G}$ with positive definite covariance matrix $\mathbf{\Sigma}_{W}$ is assumed to be independent of an arbitrary random vector $\mathbf{X}$ whose covariance matrix $\mathbf{\Sigma}_{X}$ satisfies $\mathbf{\Sigma}_{X} \preceq \mathbf{\Sigma}$. Then, there exists a Gaussian random vector $\mathbf{X}_{G}^{*}$ with covariance matrix $\mathbf{\Sigma}_{X^*}$ which satisfies the following inequality:

$$h(\mathbf{X}) - \mu h(\mathbf{X} + \mathbf{W}_{G}) \leq h(\mathbf{X}_{G}^{*}) - \mu h(\mathbf{X}_{G}^{*} + \mathbf{W}_{G}), \tag{39}$$

where $\mathbf{\Sigma}_{X^*} \preceq \mathbf{\Sigma}$.

Proof: See Appendix M. ■

Remark 5: As the extremal entropy inequality only shows the existence of necessary optimal solutions in [2] and [27], the current proof also shows the existence of necessary optimal solutions. In addition, the proposed proof only exploits calculus of variations tools. Namely, this proof does not adopt neither the channel enhancement technique and EPI in [2] nor the EPI and data processing inequality in [27].

Theorem 17: Assume that μ is an arbitrary but fixed constant, with $\mu \geq 1$, and $\mathbf{\Sigma}$ is a positive semi-definite matrix. Independent Gaussian random vectors $\mathbf{W}_{G}$ with covariance matrix $\mathbf{\Sigma}_{W}$ and $\mathbf{V}_{G}$ with covariance matrix $\mathbf{\Sigma}_{V}$ are assumed to be independent of an arbitrary random vector $\mathbf{X}$ with covariance matrix $\mathbf{\Sigma}_{X} \preceq \mathbf{\Sigma}$. Both covariance matrices $\mathbf{\Sigma}_{W}$ and $\mathbf{\Sigma}_{V}$ are assumed to be positive definite. Then, there exists a Gaussian random vector $\mathbf{X}_{G}^{*}$ with covariance matrix $\mathbf{\Sigma}_{X^*}$ which satisfies the following inequality:

$$h(\mathbf{X} + \mathbf{W}_{G}) - \mu h(\mathbf{X} + \mathbf{V}_{G}) \leq h(\mathbf{X}_{G}^{*} + \mathbf{W}_{G}) - \mu h(\mathbf{X}_{G}^{*} + \mathbf{V}_{G}), \tag{40}$$

where $\mathbf{\Sigma}_{X^*} \preceq \mathbf{\Sigma}$.

Proof: See Appendix N. ■

Remark 6: The proposed proof does not borrow any techniques from [2]. Even though the proposed proof adopts the equality condition for the data processing inequality, a result which was also exploited in [27], the proposed proof is different from the one in [27] from the following perspectives. First, the proposed proof uses the equality condition of the data processing inequality only once while the proof in [27] uses it twice. The proof in [2] exploited the channel enhancement technique twice, which is equivalent to using the equality condition in the data processing inequality. Second, the proposed proof does not use the moment generating function technique unlike the proof proposed in [27]; instead the current proof directly exploits a property of the conditional mutual information pertaining to a Markov chain.

VII. Applications

The importance of information theoretic inequalities such as entropy power inequality, extremal entropy inequality, etc., was already proved by several applications. For example, minimum Fisher information theorem (Cramér-Rao inequality) and maximum entropy theorem were used for developing min-max robust estimation techniques [31], [32]. EPI was first adapted to prove a lower bound on the capacity of additive noise channels by Shannon [28], and received huge interest recently [21], [29]. Also, EPI was exploited for the scalar Gaussian broadcast channel [24], the scalar quadratic Gaussian CEO problem [30], etc. The extremal entropy inequality can be used in the vector Gaussian broadcast channel [2], the distributed source coding with a single quadratic distortion constraint problem [2], and the Gaussian wire-tap channel [27], and so on. Even though these applications were traditionally addressed using the above mentioned information theoretic inequalities, one can directly approach these applications by means of variational calculus techniques. Numerous extensions of maximum entropy theorem, minimum Fisher information theorem, additive worst noise lemma, entropy power inequality and extremal entropy inequality might be envisioned within the proposed variational calculus framework by imposing various restrictions on the range of values assumed by random variables/vectors (e.g., random variables whose support is limited to a finite length interval) or on their second or higher-order moments and correlations. For example, the problem of finding the worst additive noise under a covariance constraint [20] as well as establishing multivariate extensions of Costa's entropy power inequality [34] along the lines mentioned by Liu et al. [26] and Palomar [35], [36] might be also addressed within the proposed variational framework.

VIII. Conclusions

In this paper, we derived several fundamental information theoretic inequalities using a functional analysis framework. The main benefit for employing calculus of variations for proving information theoretic inequalities is the fact that the global optimal solution is obtained from the necessary conditions for optimality without additional calculations. The summary of our contributions is the following. First, the entropy maximizing theorem and Fisher information minimizing theorem were derived under different assumptions. Second, the worst additive noise lemma was proved from the perspective of a functional problem. Third, the entropy power inequality and the extremal entropy inequality were derived using calculus of variations. Finally, applications and possible extensions that could be addressed based on the proposed results were briefly mentioned.

Appendix A
Proof of Theorem 4

Proof: To prove the inequality in (25), we first construct a functional problem as follows:

$$\min_{f_X} \quad \int f_X(x) \log f_X(x) dx, \tag{41}$$

$$\text{s. t.} \quad \int f_X(x) dx = 1, \tag{42}$$

$$\int x f_X(x) dx = \mu_X,$$

$$\int x^2 f_X(x) dx = m_X^2, \tag{43}$$

where μ_X is the first-order moment of X, and m_X represents the second-order moment of X.

Using Theorem 3, the functional problem in (41) is expressed as

$$\min_{f_X} \quad U[f_X], \tag{44}$$

where $U[f_X] = \int K(x, f_X) dx$, $K(x, f_X) = f_X(x)\left(\log f_X(x) + \alpha_0 + \alpha_1 x + \alpha_2 x^2\right)$, α_0, α_1, and α_2 are Lagrange multipliers.

The optimal density function f_{X^*} must satisfy the first-order variation condition as follows:

$$K'_{f_X} - \frac{d}{dx} K'_{f'_X}\Big|_{f_X = f_{X^*}} \quad = \quad 1 + \log f_{X^*}(x) + \alpha_0 + \alpha_1 x + \alpha_2 x^2 = 0. \tag{45}$$

Considering the constraints in (42)-(43) and the equation in (45), it follows that

$$\begin{aligned} f_{X^*}(x) &= \frac{1}{\sqrt{2\pi \frac{1}{2\alpha_2}}} \exp\left\{ -\frac{1}{2\frac{1}{2\alpha_2}} \left(x + \frac{\alpha_1}{2\alpha_2} \right)^2 \right\} \sqrt{2\pi \frac{1}{2\alpha_2}} \exp\left\{ -\alpha_0 - 1 + \frac{\alpha_1^2}{4\alpha_2} \right\} \\ &= \frac{1}{\sqrt{2\pi (m_X^2 - \mu_X^2)}} \exp\left\{ -\frac{1}{2(m_X^2 - \mu_X^2)} (x - \mu_X)^2 \right\}, \end{aligned} \tag{46}$$

where

$$\begin{aligned} \alpha_0 &= -1 + \frac{\mu_X^2}{2(m_X^2 - \mu_X^2)} + \frac{1}{2} \log 2\pi (m_X^2 - \mu_X^2), \\ \alpha_1 &= -\frac{\mu_X}{m_X^2 - \mu_X^2}, \\ \alpha_2 &= \frac{1}{2(m_X^2 - \mu_X^2)}. \end{aligned} \tag{47}$$

Since the second-order variation of $U[f_X]$ is expressed as

$$K''_{f_X f_X}\Big|_{f_X = f_{X^*}} \quad = \quad \frac{1}{f_{X^*}(x)}, \tag{48}$$

and it is positive, the optimal solution f_{X^*} minimizes the variational problem in (41).

These first-order and second-order conditions are not sufficient but necessary for the optimal solution. However, as shown in (45) and (46), there exists only one solution, the Gaussian density function, in the feasible set. Therefore, the Gaussian density function is also sufficient in this case.

Therefore, a negative differential entropy $-h(X)$ is minimized (or, equivalently $h(X)$ is maximized) when $f_X(x)$ is Gaussian, and the proof is completed. ■

Appendix B
Proof of Theorem 5

Proof: We first construct a functional problem, which represents the inequality in (26) and required constraints, as follows:

$$\min_{f_X} \quad \int f_X(\mathbf{x}) \log f_X(\mathbf{x}) d\mathbf{x}, \tag{49}$$

$$\text{s. t.} \quad \int f_X(\mathbf{x}) d\mathbf{x} = 1, \tag{50}$$

$$\int \mathbf{x}\mathbf{x}^T f_X(\mathbf{x}) d\mathbf{x} = \mathbf{\Omega}_X. \tag{51}$$

Using Theorem 3, the functional problem in (49) is expressed as

$$\min_{f_X} \quad U[f_X], \tag{52}$$

where $U[f_X] = \int K(\mathbf{x}, f_X) d\mathbf{x} = \int f_X(\mathbf{x}) \left(\log f_X(\mathbf{x}) + \alpha + \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_{ij} x_i x_j \right) d\mathbf{x}$, and α and λ_{ij} are Lagrange multipliers.

Based on Theorem 1 or Corollary 2, by checking the first-order variation condition, we can find the optimal solution $f_{X^*}(\mathbf{x})$ as follows.

$$K'_{f_X} - \frac{d}{dx} K'_{f'_X} \Big|_{f_X = f_{X^*}} = 1 + \log f_{X^*}(\mathbf{x}) + \alpha + \mathbf{x}^T \mathbf{\Lambda} \mathbf{x} = 0, \tag{53}$$

$$\tag{54}$$

Considering the constraints in (50) and (51),

$$\begin{aligned} f_{X^*}(\mathbf{x}) &= \exp\{-\mathbf{x}^T \mathbf{\Lambda} \mathbf{x} - \alpha - 1\} \\ &= (2\pi)^{-\frac{n}{2}} \left| \frac{1}{2} \mathbf{\Lambda}^{-1} \right|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} \mathbf{x}^T \left(\frac{1}{2} \mathbf{\Lambda}^{-1} \right)^{-1} \mathbf{x} \right\} (2\pi)^{\frac{n}{2}} \left| \frac{1}{2} \mathbf{\Lambda}^{-1} \right|^{\frac{1}{2}} \exp\{-1 - \alpha\} \\ &= (2\pi)^{-\frac{n}{2}} |\mathbf{\Omega}_X|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} \mathbf{x}^T \mathbf{\Omega}_X^{-1} \mathbf{x} \right\}, \end{aligned} \tag{55}$$

where

$$\begin{aligned}\alpha &= -1+\frac{1}{2}\log\left(2\pi\right)^{n}\left|\mathbf{\Omega}_{X}\right|,\\ \mathbf{\Lambda} &= \frac{1}{2}\mathbf{\Omega}_{X}^{-1}.\end{aligned} \tag{56}$$

Here, two remarks are in order. First, the correlation matrix $\mathbf{\Omega}_X$ is assumed to be invertible. When the correlation matrix is non-invertible, similar to the method shown in [2], we can equivalently re-write the functional problem in (49) and its constraints in (51) as

$$\min_{f_{\bar{X}}} \quad \int f_{\bar{X}}(\mathbf{x})\log f_{\bar{X}}(\mathbf{x})d\mathbf{x}, \tag{57}$$

$$\begin{aligned}\text{s. t.} \quad & \int f_{\bar{X}}(\mathbf{x})d\mathbf{x}=1,\\ & \int \mathbf{x}\mathbf{x}^{T} f_{\bar{X}}(\mathbf{x})d\mathbf{x}=\mathbf{\Omega}_{\bar{X}},\end{aligned} \tag{58}$$

where $\bar{\mathbf{X}}$ is a random vector with correlation matrix $\mathbf{\Omega}_{\bar{X}}$, and $\mathbf{\Omega}_{\bar{X}}$ is a positive definite matrix. Therefore, without loss of generality, we assume the correlation matrix $\mathbf{\Omega}_X$ is invertible. Second, if an additional constraint, related to the mean vector of $\mathbf{X}$, $\boldsymbol{\mu}_X$, is given, the optimal solution is a multi-variate Gaussian density function, whose mean is $\boldsymbol{\mu}_X$, instead of the multi-variate Gaussian density function, which has zero mean, in (55) (cf. Appendix A).

Since

$$K''_{f_X f_X}\Big|_{f_X=f_{X^*}} = \frac{1}{f_{X^*}(\mathbf{x})} > 0,$$

the second-order variation $\delta^2 U\left[f_{X^*}\right]$ is positive, and the optimal solution f_{X^*} is a minimal solution for the variational problem in (49).

Therefore, a differential entropy $-h(\mathbf{X})$ is minimized (or, equivalently $h(\mathbf{X})$ is maximized) when $\mathbf{X}$ is a multi-variate Gaussian random vector with zero mean and a covariance matrix $\mathbf{\Sigma}_X$. Even though Theorems 1, 2 are necessary conditions for the minimum, in this case, a multi-variate Gaussian density function is an actual solution since there is only one solution, a multi-variate Gaussian density function, in the feasible set. ■

Appendix C
Proof of Theorem 6

Proof: We first construct a functional problem, which represents the inequality in (27) and required constraints, as follows:

$$\begin{aligned} \min_{f_X} \quad & \int_0^\infty f_X(x) \log f_X(x) dx, & (59) \\ \text{s. t.} \quad & \int_0^\infty f_X(x) dx = 1, \\ & \int_0^\infty x^2 f_X(x) dx = m_X^2. & (60) \end{aligned}$$

Using Theorem 3, the functional problem in (59) is expressed as

$$\min_{f_X} \quad U[f_X], \tag{61}$$

where $U[f_X] = \int K(x, f_X) dx$, $K(x, f_X) = f_X(x) \left(\log f_X(x) + \alpha_0 + \alpha_1 x^2 \right)$, and α_0 and α_1 are Lagrange multipliers.[2]

Based on Theorem 1 or Corollary 2, the first-order variation condition of $U[f_X]$ is considered as follows.

$$K'_{f_X} - \frac{d}{dx} K'_{f'_X} \Big|_{f_X = f_{X^*}} = 1 + \log f_{X^*}(x) + \alpha_0 + \alpha_1 x^2 = 0. \tag{62}$$

Considering the constraints in (60) and the equation in (62),

$$\begin{aligned} f_{X^*}(x) &= \frac{1}{\sqrt{\pi \frac{1}{4\alpha_1}}} \exp\left\{ -\frac{1}{2 \frac{1}{2\alpha_1}} x^2 \right\} \sqrt{\pi \frac{1}{4\alpha_1}} \exp\{-\alpha_0 - 1\} \\ &= \frac{1}{\sqrt{\frac{\pi m_X^2}{2}}} \exp\left\{ -\frac{1}{2 m_X^2} x^2 \right\}, \qquad x \geq 0, \end{aligned} \tag{63}$$

where

$$\begin{aligned} \alpha_0 &= -1 + \frac{1}{2} \log \frac{\pi m_X^2}{2}, \\ \alpha_1 &= \frac{1}{2 m_X^2}. \end{aligned}$$

Since

$$K''_{f_X f_X} \Big|_{f_X = f_{X^*}} = \frac{1}{f_{X^*}(x)} > 0,$$

[2] For the simplicity of notations, the range of integration will not be explicitly expressed in the rest of this proof. Throughout the paper, the range of integration will not be explicitly denoted unless the range is ambiguous.

and the second-order variation $\delta^2 U[f_{X^*}] > 0$, the optimal solution f_{X^*} is a minimal solution for the variational problem in (59).

These first-order and second-order conditions are not sufficient but necessary for the optimal. However, as shown in (62) and (63), there exists only one solution, a half-normal density function, in the feasible set. Therefore, a half-normal density function is also sufficient in this problem.

Therefore, given the second-order moment, the negative differential entropy $-h(X)$ is minimized (or, equivalently $h(X)$ is maximized) over the set of non-negative random variables when $f_X(x)$ is a half-normal density function.

Remark 7: Since a half-normal random variable has a fixed mean, if we add a constraint of the mean such as $\mathbb{E}_X[X] = \mu_X$ in (60), the inequality in (27) is not true except when $\mu_X = \sqrt{2m_X^2/\pi}$, where μ_X and m_X^2 are the first-order moment and the second-order moment of X, respectively.

■

APPENDIX D
PROOF OF THEOREM 7

Proof: We first construct a functional problem, which represents the inequality in (28) and required constraints, as follows:

$$\begin{aligned}\min_{f_X} \quad & \int \frac{f_X'(x)^2}{f_X(x)}dx, && (64)\\ \text{s. t.} \quad & \int f_X(x)dx = 1, \\ & \int x f_X(x)dx = \mu_X, \\ & \int x^2 f_X(x)dx = m_X^2. && (65)\end{aligned}$$

Using Theorem 3, the functional problem in (64) is expressed as

$$\min_{f_X} \quad U[f_X], \tag{66}$$

where $U[f_X] = \int K(x, f_X, f_X')dx$, $K(x, f_X, f_X') = (f_X'(x)^2/f_X(x)) + f_X(x)\left(\alpha_0 + \alpha_1 x + \alpha_2 x^2\right)$, and α_0, α_1, and α_2 are the Lagrange multipliers.

Based on Theorem 1 or Corollary 2, the first-order variation is investigated as follows:

$$K_{f_X}' - \frac{d}{dx}K_{f_X'}'\Big|_{f_X = f_{X^*}} = \left(\frac{f_{X^*}'(x)}{f_{X^*}(x)}\right)^2 - 2\frac{f_{X^*}''(x)}{f_{X^*}(x)} + \alpha_0 + \alpha_1 x + \alpha_2 x^2 = 0. \tag{67}$$

Unlike Theorem 4, we cannot directly calculate $f_{X^*}(x)$ from the equation in (67). Fortunately, when $f_{X^*}(x)$ is a Gaussian density function, $(f'_{X^*}(x)/f_{X^*}(x))^2 - 2(f''_{X^*}(x)/f_{X^*}(x))$ in (67) is expressed as a quadratic function, which is similar to the quadratic parts in (67).

Due to the constraints in (65), a Gaussian density function $f_{X^*}(x)$ is defined as

$$f_{X^*}(x) = \frac{1}{\sqrt{2\pi(m_x^2 - \mu_x^2)}} \exp\left\{-\frac{1}{2\left(m_x^2 - \mu_x^2\right)}\left(x - \mu_X\right)^2\right\}. \tag{68}$$

By substituting $f_{X^*}(x)$ in (68) for the equation in (67),

$$\begin{aligned} & \left(-\frac{1}{m_x^2 - \mu_x^2}\left(x - \mu_X\right)\right)^2 - 2\left\{\left(-\frac{1}{m_x^2 - \mu_x^2}\left(x - \mu_X\right)\right)^2 - \frac{1}{m_x^2 - \mu_x^2}\right\} + \alpha_0 + \alpha_1 x + \alpha_2 x^2 \\ = \;& -\frac{1}{(m_x^2 - \mu_x^2)^2}x^2 + \frac{2\mu_X}{(m_x^2 - \mu_x^2)^2}x + \left(-\frac{\mu_x^2}{(m_x^2 - \mu_x^2)^2} + \frac{2}{m_x^2 - \mu_x^2}\right) + \alpha_0 + \alpha_1 x + \alpha_2 x^2 \\ = \;& 0. \end{aligned} \tag{69}$$

Since the equations in (69) must be satisfied for any x,

$$\begin{aligned} \alpha_0 &= \frac{\mu_x^2}{(m_x^2 - \mu_x^2)^2} - \frac{2}{m_x^2 - \mu_x^2}, \\ \alpha_1 &= -\frac{2\mu_X}{(m_x^2 - \mu_x^2)^2}, \\ \alpha_2 &= \frac{1}{(m_x^2 - \mu_x^2)^2}. \end{aligned} \tag{70}$$

Since

$$K''_{f'_{X^*}f'_{X^*}} = 2\frac{1}{f_{X^*}(x)} > 0 \tag{71}$$

and the second-order variation $\delta^2 U[f_{X^*}]$ is positive, the optimal solution f_{X^*} minimizes the variational problem in (64).

Therefore, Fisher information $J(X)$ is minimized when $f_X(x)$ is Gaussian. Even though Theorems 1, 2 are necessary conditions for the minimum, in this case, a Gaussian density function is sufficiently optimal due to the following fact: the objective function is strictly convex and the constraint sets are convex. Therefore, the proof is completed.

Remark 8: Even though this result is well-known in the literature (e.g., [18], [21]), this is the first rigorous proof based on calculus of variations.

Remark 9: The constraint related to the first-order moment in (65), is not required in this case. Without the constraint, the optimal solution is a Gaussian density function, which has zero mean.

■

APPENDIX E
PROOF OF THEOREM 8

Proof: We first construct a functional problem, which represents the inequality in (29) and the required constraints as follows:

$$\begin{aligned} \min_{f_X} \quad & \int \boldsymbol{\xi}^T \nabla f_X(\mathbf{x}) \nabla f_X(\mathbf{x})^T \boldsymbol{\xi} \frac{1}{f_X(\mathbf{x})} d\mathbf{x}, & (72) \\ \text{s. t.} \quad & \int f_X(\mathbf{x}) d\mathbf{x} = 1, \\ & \int \mathbf{x} f_X(\mathbf{x}) d\mathbf{x} = \boldsymbol{\mu}_X, \\ & \int \mathbf{x}\mathbf{x}^T f_X(\mathbf{x}) d\mathbf{x} = \boldsymbol{\Omega}_X, & (73) \end{aligned}$$

where $\boldsymbol{\xi}$ is an arbitrary but fixed non-zero vector, and it is defined as $\boldsymbol{\xi} = [\xi_1, \ldots, \xi_n]^T$.

Using Theorem 3, the functional problem in (72) is expressed as

$$\min_{f_X} \quad U[f_X], \tag{74}$$

where $U[f_X] = \int K(\mathbf{x}, f_X, \nabla f_X) d\mathbf{x}$, $K(\mathbf{x}, f_X, \nabla f_X) = (\boldsymbol{\xi}^T \nabla f_X(\mathbf{x}) \nabla f_X(\mathbf{x})^T \boldsymbol{\xi} / f_X(\mathbf{x})) + f_X(\mathbf{x}) \sum_{i=1}^n \zeta_i x_i + \alpha f_X(\mathbf{x}) + f_X(\mathbf{x}) \sum_{i=1}^n \sum_{j=1}^n \lambda_{ij} x_i x_j$, and α, ζ_i, and λ_{ij} are Lagrange multipliers.

Based on Theorem 1 or 2, by confirming the first-order variation condition, i.e., $\delta U[f_{X^*}] = 0$, we can find the optimal solution $f_{X^*}(x)$ as follows.

$$K'_{f_X} - \sum_{i=1}^n \frac{\partial}{\partial x_i} K'_{f'_{X_i}} \bigg|_{f_X = f_{X^*}} = 0, \tag{75}$$

where

$$\begin{aligned} K'_{f_X} &= -\frac{\boldsymbol{\xi}^T \nabla f_X(\mathbf{x}) \nabla f_X(\mathbf{x})^T \boldsymbol{\xi}}{f_X(\mathbf{x})^2} + \alpha + \boldsymbol{\zeta}^T \mathbf{x} + \mathbf{x}^T \boldsymbol{\Lambda} \mathbf{x}, \\ \frac{\partial}{\partial x_i} K'_{f'_{X_i}} &= \frac{\partial}{\partial x_i} \left(\frac{2 \sum_{j=1}^n \frac{\partial}{\partial x_j} f_X(\mathbf{x}) \xi_i \xi_j}{f_X(\mathbf{x})} \right) \\ &= \frac{2 \sum_{j=1}^n \frac{\partial}{\partial x_i} \frac{\partial}{\partial x_j} f_X(\mathbf{x}) \xi_i \xi_j}{f_X(\mathbf{x})} - \frac{2 \sum_{j=1}^n \frac{\partial}{\partial x_j} f_X(\mathbf{x}) \xi_i \xi_j \frac{\partial}{\partial x_i} f_X(\mathbf{x})}{f_X(\mathbf{x})^2}. \end{aligned} \tag{76}$$

Therefore, the left-hand side of the equation in (75) is expressed as

$$\begin{aligned}
& K'_{f_X} - \sum_{i=1}^{n} \frac{\partial}{\partial x_i} K'_{f_{X_i}} \\
= \quad & \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial}{\partial x_i} f_X(\mathbf{x}) \frac{\partial}{\partial x_j} f_X(\mathbf{x}) \xi_i \xi_j}{f_X(\mathbf{x})^2} - \frac{2\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial}{\partial x_i}\frac{\partial}{\partial x_j} f_X(\mathbf{x}) \xi_i \xi_j}{f_X(\mathbf{x})} + \alpha + \sum_{i=1}^{n} \zeta_i x_i + \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_{ij} x_i x_j \qquad (77) \\
= \quad & 0. \qquad (78)
\end{aligned}$$

Unlike Theorem 5, we cannot directly calculate $f_{X^*}(\mathbf{x})$ from the equation in (75). Fortunately, the first two parts in equation (77) are expressed as a quadratic function when $f_{X^*}(\mathbf{x})$ is a multi-variate Gaussian density function, and therefore, the multi-variate Gaussian density function satisfies the equality in (78). When $f_{X^*}(\mathbf{x})$ is a multi-variate Gaussian density function:

$$f_{X^*}(\mathbf{x}) = (2\pi)^{-\frac{n}{2}} \left|\mathbf{\Sigma}_X\right|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left(\mathbf{x}-\boldsymbol{\mu}_X\right)^T \mathbf{\Sigma}_X^{-1}\left(\mathbf{x}-\boldsymbol{\mu}_X\right)\right\},$$

where $\mathbf{\Sigma}_X = \mathbf{\Omega}_X - \boldsymbol{\mu}_X \boldsymbol{\mu}_X^T$,

$$\mathbf{\Sigma}_X^{-1} = \begin{bmatrix} \sigma^2_{X_{11}} & \cdots & \sigma^2_{X_{1n}} \\ \vdots & \ddots & \vdots \\ \sigma^2_{X_{n1}} & \cdots & \sigma^2_{X_{nn}} \end{bmatrix}, \qquad (79)$$

its partial derivatives are expressed as follows:

$$\begin{aligned}
\frac{\partial}{\partial x_i} f_{X^*}(\mathbf{x}) &= -\frac{1}{2}\left(\sum_{l=1}^{n} \sigma^2_{X_{il}}\left(x_l - \mu_{X_l}\right) + \sum_{m=1}^{n} \sigma^2_{X_{mi}}\left(x_m - \mu_{X_m}\right)\right) f_{X^*}(\mathbf{x}) \\
\frac{\partial}{\partial x_j}\frac{\partial}{\partial x_i} f_{X^*}(\mathbf{x}) &= -\frac{1}{2}\left(\sigma^2_{X_{ij}} + \sigma^2_{X_{ji}}\right) f_{X^*}(\mathbf{x}) + \frac{1}{4}\left(\sum_{l=1}^{n} \sigma^2_{X_{il}}\left(x_l - \mu_{X_l}\right) + \sum_{m=1}^{n} \sigma^2_{X_{mi}}\left(x_m - \mu_{X_m}\right)\right) \\
&\qquad \times \left(\sum_{l=1}^{n} \sigma^2_{X_{jl}}\left(x_l - \mu_{X_l}\right) + \sum_{m=1}^{n} \sigma^2_{X_{mj}}\left(x_m - \mu_{X_m}\right)\right) f_{X^*}(\mathbf{x}). \qquad (80)
\end{aligned}$$

Without loss of generality, the covariance matrix $\mathbf{\Sigma}_X$ is assumed to be invertible due to the same reason mentioned in Appendix B.

By substituting the equations in (80) into the equations (77), it turns out that

$$
\begin{aligned}
& K'_{f_{X^*}} - \sum_{i=1}^{n} \frac{\partial}{\partial x_i} K'_{f'_{X_i^*}} \\
= \; & \frac{1}{4} \sum_{i=1}^{n} \sum_{j=1}^{n} \xi_i \xi_j \left(\sum_{l=1}^{n} \left(\sigma^2_{X_{il}} + \sigma^2_{X_{li}} \right) \left(x_l - \mu_{X_l} \right) \right) \left(\sum_{m=1}^{n} \left(\sigma^2_{X_{jm}} + \sigma^2_{X_{mj}} \right) \left(x_m - \mu_{X_m} \right) \right) \\
& + \sum_{i=1}^{n} \sum_{j=1}^{n} \left(\sigma^2_{X_{ij}} + \sigma^2_{X_{ji}} \right) \xi_i \xi_j + \alpha + \sum_{i=1}^{n} \zeta_i x_i + \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_{ij} x_i x_j \\
= \; & \sum_{l=1}^{n} \sum_{m=1}^{n} \left[\left(x_l - \mu_{X_l} \right) \left(x_m - \mu_{X_m} \right) \left(\frac{1}{4} \sum_{i=1}^{n} \sum_{j=1}^{n} \xi_i \xi_j \left(\sigma^2_{X_{il}} + \sigma^2_{X_{li}} \right) \left(\sigma^2_{X_{jm}} + \sigma^2_{X_{mj}} \right) \right) \right] \\
& + \sum_{i=1}^{n} \sum_{j=1}^{n} \left(\sigma^2_{X_{ij}} + \sigma^2_{X_{ji}} \right) \xi_i \xi_j + \alpha + \sum_{i=1}^{n} \zeta_i x_i + \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_{ij} x_i x_j \\
= \; & \sum_{l=1}^{n} \sum_{m=1}^{n} \left[\left(x_l - \mu_{X_l} \right) \left(x_m - \mu_{X_m} \right) \boldsymbol{\xi}^T \boldsymbol{\Sigma}_{X_{lm}} \boldsymbol{\xi} \right] + \sum_{i=1}^{n} \sum_{j=1}^{n} \left(\sigma^2_{X_{ij}} + \sigma^2_{X_{ji}} \right) \xi_i \xi_j \\
& + \alpha + \sum_{i=1}^{n} \zeta_i x_i + \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_{ij} x_i x_j \\
= \; & \sum_{l=1}^{n} \sum_{m=1}^{n} \omega_{lm} \left(x_l - \mu_{X_l} \right) \left(x_m - \mu_{X_m} \right) + \sum_{i=1}^{n} \sum_{j=1}^{n} \left(\sigma^2_{X_{ij}} + \sigma^2_{X_{ji}} \right) \xi_i \xi_j + \alpha + \sum_{i=1}^{n} \zeta_i x_i + \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_{ij} x_i x_j \\
= \; & \left(\mathbf{x} - \boldsymbol{\mu}_X \right)^T \boldsymbol{\Omega} \left(\mathbf{x} - \boldsymbol{\mu}_X \right) + \boldsymbol{\xi}^T \boldsymbol{\Psi} \boldsymbol{\xi} + \alpha + \boldsymbol{\zeta}^T \mathbf{x} + \mathbf{x}^T \boldsymbol{\Lambda} \mathbf{x} \\
= \; & \left(\mathbf{x}^T \boldsymbol{\Omega} \mathbf{x} + \mathbf{x}^T \boldsymbol{\Lambda} \mathbf{x} \right) + \left(\boldsymbol{\zeta}^T \mathbf{x} - 2 \boldsymbol{\mu}_X^T \boldsymbol{\Omega} \mathbf{x} \right) + \boldsymbol{\mu}_X^T \boldsymbol{\Omega} \boldsymbol{\mu}_X + \boldsymbol{\xi}^T \boldsymbol{\Psi} \boldsymbol{\xi} + \alpha \\
= \; & 0,
\end{aligned} \tag{81}
$$

where

$$\begin{aligned}
\boldsymbol{\Sigma}_{X_{lm}} &= \begin{bmatrix} \Sigma^{lm}_{X_{11}} & \cdots & \Sigma^{lm}_{X_{1n}} \\ \vdots & \ddots & \vdots \\ \Sigma^{lm}_{X_{n1}} & \cdots & \Sigma^{lm}_{X_{nn}} \end{bmatrix}, \quad \boldsymbol{\Lambda} = \begin{bmatrix} \lambda_{11} & \cdots & \lambda_{1n} \\ \vdots & \ddots & \vdots \\ \lambda_{n1} & \cdots & \lambda_{nn} \end{bmatrix}, \\
\boldsymbol{\Psi} &= \begin{bmatrix} \psi_{11} & \cdots & \psi_{1n} \\ \vdots & \ddots & \vdots \\ \psi_{n1} & \cdots & \psi_{nn} \end{bmatrix}, \quad \boldsymbol{\Omega} = \begin{bmatrix} \omega_{11} & \cdots & \omega_{1n} \\ \vdots & \ddots & \vdots \\ \omega_{n1} & \cdots & \omega_{nn} \end{bmatrix}, \\
\Sigma^{lm}_{X_{ij}} &= \frac{1}{4}\left(\sigma^2_{X_{il}} + \sigma^2_{X_{li}}\right)\left(\sigma^2_{X_{jm}} + \sigma^2_{X_{mj}}\right) \\
&= \sigma^2_{X_{li}}\sigma^2_{X_{jm}}, \qquad i = 1, \ldots, n, \quad j = 1, \ldots, n, \quad l = 1, \ldots, n, \quad m = 1, \ldots, n, \\
\psi_{ij} &= 2\sigma^2_{X_{ij}}, \qquad i = 1, \ldots, n, \quad j = 1, \ldots, n, \\
\omega_{lm} &= \boldsymbol{\xi}^T \boldsymbol{\Sigma}_{X_{lm}} \boldsymbol{\xi}, \qquad l = 1, \ldots, n, \quad m = 1, \ldots, n.
\end{aligned} \tag{82}$$

Therefore, the Lagrange multipliers α and λ_{ij} are defined as

$$\begin{aligned}
\alpha &= -\boldsymbol{\mu}_X^T \boldsymbol{\Omega} \boldsymbol{\mu}_X - \boldsymbol{\xi}^T \boldsymbol{\Psi} \boldsymbol{\xi}, \\
\boldsymbol{\zeta} &= 2\boldsymbol{\Omega}\boldsymbol{\mu}_X, \\
\boldsymbol{\Lambda} &= -\boldsymbol{\Omega}.
\end{aligned} \tag{83}$$

Since the second-order variation condition is positive

$$K''_{f'_X f'_X} = 2\frac{1}{f_{X^*}(\mathbf{x})} > 0, \tag{84}$$

the optimal solution $f_{X^*}(\mathbf{x})$ minimizes the variational problem in (72). Therefore, the Fisher information matrix $\mathbb{J}(\mathbf{X})$ is minimized when $f_{X^*}(\mathbf{x})$ is a multi-variate Gaussian, i.e., $\mathbb{J}(\mathbf{X}) \succeq \mathbb{J}(\mathbf{X}_G)$. Even though Theorems 1, 2 are necessary conditions for the minimum, in this case, the multi-variate Gaussian density function is sufficiently the global minimum solution since the objective function is strictly convex and its constraint sets are convex. ■

APPENDIX F
PROOF OF THEOREM 9

Proof: We first construct a functional problem, which represents the inequality in (31) and required constraints, as follows:

$$\begin{aligned} \min_{f_X} \quad & \int_0^\infty \frac{f_X'(x)^2}{f_X(x)}dx, & (85) \\ \text{s. t.} \quad & \int_0^\infty f_X(x)dx = 1, \\ & \int_0^\infty x^2 f_X(x)dx = m_X^2. & (86) \end{aligned}$$

Using Theorem 3, the functional problem in (85) is expressed as

$$\min_{f_X} \quad U[f_X], \tag{87}$$

where $U[f_X] = \int K(x, f_X, f_X')dx$, $K(x, f_X, f_X') = (f_X'(x)^2/f_X(x)) + f_X(x)\left(\alpha_0 + \alpha_1 x^2\right)$, and α_0 and α_1 are the Lagrange multipliers.

Based on Theorem 1 or 2, the first-order and the second-order variation conditions of $U[f_X]$ will be considered as follows. First, the optimal solution $f_{X^*}(x)$ must satisfy the following first-order variation condition:

$$K_{f_X}' - \frac{d}{dx}K_{f_X'}'\Big|_{f_X = f_{X^*}} = \left(\frac{f_{X^*}'(x)}{f_{X^*}(x)}\right)^2 - 2\frac{f_{X^*}''(x)}{f_{X^*}(x)} + \alpha_0 + \alpha_1 x^2 = 0. \tag{88}$$

When $f_{X^*}(x)$ is a half-normal density function, $(f_{X^*}'(x)/f_{X^*}(x))^2 - 2(f_{X^*}''(x)/f_{X^*}(x))$ in (88) is expressed as a quadratic function, and therefore the equation in (88) can be satisfied.

Considering the constraints in (86) and $f_{X^*}(x) = (1/\sqrt{\pi m_X^2/2})\exp(-x^2/(2m_X^2)))$, where $x > 0$,

$$\begin{aligned} & \left(-\frac{1}{m_X^2}x\right)^2 - 2\left\{\left(-\frac{1}{m_X^2}x\right)^2 - \frac{1}{m_X^2}\right\} + \alpha_0 + \alpha_1 x^2 \\ = \; & -\frac{1}{m_X^4}x^2 + \frac{2}{m_X^2} + \alpha_0 + \alpha_1 x^2 \\ = \; & 0. \end{aligned} \tag{89}$$

Since the equation in (89) is satisfied for any x,

$$\begin{aligned} \alpha_0 &= -\frac{2}{m_X^2}, \\ \alpha_1 &= \frac{1}{m_X^4}. \end{aligned} \tag{90}$$

Now, the second-order variation condition is considered as follows. Since

$$K''_{f'_X f'_X}\Big|_{f_X=f_{X^*}} \quad = \quad 2\frac{1}{f_{X^*}(x)} > 0, \tag{91}$$

the second-order variation of $\delta^2 U[f_{X^*}] > 0$, and therefore f_{X^*} minimizes the variational problem in (33). Therefore, the Fisher information $J(X)$ is minimized when $f_X(x)$ is half normal. Even though Theorems 1, 2 are necessary conditions for the minimum, in this case, a half normal density function is sufficiently optimal due to the strict convexity of the objective function and the convexity of the constraint set in (85) and (86). Therefore, the proof is completed. ■

APPENDIX G

PROOF OF THEOREM 10

Proof: We first construct a functional problem, which represents the inequality in (33) and required constraints, as follows:

$$\min_{f_X} \quad \int \frac{f'_X(x)^2}{f_X(x)} dx, \tag{92}$$

$$\text{s. t.} \quad \int f_X(x) dx = 1,$$

$$\int x^2 f_X(x) dx = m_X^2. \tag{93}$$

Using Theorem 3, the functional problem in (92) is expressed as

$$\min_{f_X} \quad U[f_X], \tag{94}$$

where $U[f_X] = \int K(x, f_X, f'_X)dx$, $K(x, f_X, f'_X) = (f'_X(x)^2/f_X(x)) + f_X(x)\left(\alpha_0 + \alpha_1 x^2\right)$, and α_0 and α_1 are the Lagrange multipliers.

Based on Theorem 1 or Corollary 2, by confirming the first-order variation condition, the optimal solution $f_{X^*}(x)$ can be found as follows:

$$K'_{f_X} - \frac{d}{dx}K'_{f'_X}\Big|_{f_X=f_{X^*}} \quad = \quad \left(\frac{f'_{X^*}(x)}{f_{X^*}(x)}\right)^2 - 2\frac{f''_{X^*}(x)}{f_{X^*}(x)} + \alpha_0 + \alpha_1 x^2 = 0. \tag{95}$$

Unfortunately, we cannot directly calculate $f_{X^*}(x)$ from the equation in (95). Instead, we try to search density functions which satisfy the equation in (95). The first two parts, $(f'_{X^*}(x)/f_{X^*}(x))^2 - 2(f''_{X^*}(x)/f_{X^*}(x))$, in equation (95) are expressed as a quadratic function when $f_{X^*}(x)$ is a chi density function with 3 degrees of freedom. Therefore, the chi density function satisfies the equation in (95).

Considering the constraints in (93) and defining $f_{X^*}(x)$ as $\sqrt{2/\pi}a^{-3}x^2\exp(-x^2/(2a^2))$, where $a = \sqrt{m_X^2/3}$, the equation in (95) is expressed as

$$\begin{aligned} & \left(\frac{2}{x} - \frac{x}{a^2}\right)^2 - 2\left(\frac{x^2}{a^4} + \frac{2}{x^2} - \frac{5}{a^2}\right) + \alpha_0 + \alpha_1 x^2 \\ = \;& -\frac{1}{a^4}x^2 + \frac{6}{a^2} + \alpha_0 + \alpha_1 x^2 \\ = \;& 0. \end{aligned} \tag{96}$$

Since the equation in (96) must be satisfied for any x,

$$\begin{aligned} \alpha_0 &= -\frac{6}{a^2} = -\frac{18}{m_X^2}, \\ \alpha_1 &= \frac{1}{a^4} = \left(\frac{3}{m_X^2}\right)^2. \end{aligned} \tag{97}$$

Now, using the second-order variation condition, we will confirm that the optimal solution f_{X^*} actually minimizes the variational problem in (92) as shown in the following equation:

$$K''_{f'_X f'_X}\Big|_{f_X = f_{X^*}} = 2\frac{1}{f_{X^*}(x)} > 0. \tag{98}$$

Therefore, the Fisher information $J(X)$ is minimized when $f_X(x)$ is a chi density function with 3 degrees of freedom and the second-order moment m_X^2. Even though Theorems 1, 2 are necessary conditions for the minimum, in this case, the chi density function is sufficiently minimum since the variational problem in (92) is strictly convex and the constraint set in (93) is convex. Therefore, the proof is completed.

Remark 10: Both a half normal density function and a chi-density function satisfy Euler's equation. Therefore, these two functions are the optimal solutions which minimize Fisher information for non-negative random variables. However, a half normal density function does not obey the regularity condition for Fisher information while a chi density function satisfies the regularity condition.

■

APPENDIX H

PROOF OF THEOREM 11

Proof: To prove the inequality in (34), the following functional problem is constructed:

$$\min_{f_X} \int\int f_X(x) f_{Y|X}(y|x) \left[-\log\left(\int f_X(x) f_{Y|X}(y|x) dx \right) + \log f_X(x) \right] dxdy \quad (99)$$

$$\text{s. t.} \quad \int f_X(x) dx = 1,$$
$$\int x^2 f_X(x) dx = m_X^2. \quad (100)$$

After substituting the random variable Y for $X + W_G$, its density function $f_Y(y)$ is expressed as

$$\begin{aligned} f_Y(y) &= \int f_X(x) f_{Y|X}(y|x) dx \\ &= \int f_X(x) f_W(y-x) dx. \end{aligned} \quad (101)$$

Then, the problem in (99) and its constraints in (100) are expressed as

$$\min_{f_X, f_Y} \int\int f_X(x) f_W(y-x) \left[-\log f_Y(y) + \log f_X(x) \right] dxdy \quad (102)$$

$$\begin{aligned} \text{s. t.} \quad & \int\int f_X(x) f_W(y-x) dxdy = 1, \\ & \int\int x^2 f_X(x) f_W(y-x) dxdy = m_X^2, \\ & \int y^2 f_Y(x) dy = m_Y^2, \\ & f_Y(y) = \int f_X(x) f_W(y-x) dx. \end{aligned} \quad (103)$$

Using Lagrange multipliers, the functional problem in (102) is denoted as

$$\min_{f_X, f_Y} \int \left(\int f_X(x) f_W(y-x) \left[-\log f_Y(y) + \log f_X(x) + \alpha_0 + \alpha_1 x^2 - \lambda(y) \right] dx \right.$$
$$\left. + f_Y(y) \left[\alpha_2 y^2 + \lambda(y) \right] \right) dy. \quad (104)$$

Define a functional U as

$$U[f_X, f_Y] = \int \left(\int K(x, y, f_X, f_Y) dx \right) + \tilde{K}(y, f_Y) dy, \quad (105)$$

where[3] $K(x, y, f_X, f_Y) = f_X(x) f_W(y-x)[-\log f_Y(y) + \log f_X(x) + \alpha_0 + \alpha_1 x^2 - \lambda(y)]$, and $\tilde{K}(y, f_Y) = f_Y(y) \left[\alpha_2 y^2 + \lambda(y) \right]$.

[3]The equation in (105) is denoted as $\int(\int K dx) + \tilde{K} dy$ for the simplicity of notation.

Now, we have to find f_{X^*} and f_{Y^*} which satisfy the first-order variation condition, $\delta U = 0$.

$$\begin{aligned} K'_{f_X}\Big|_{f_X=f_{X^*},f_Y=f_{Y^*}} &= f_W(y-x)\left(-\log f_{Y^*}(y)+\log f_{X^*}(x)+\alpha_0+\alpha_1 x^2+1-\lambda(y)\right) \\ &= 0 \end{aligned} \tag{106}$$

$$\begin{aligned} \int K'_{f_Y}dx+\tilde{K}'_{f_Y}\Big|_{f_X=f_{X^*},f_Y=f_{Y^*}} &= -\int f_{X^*}(x)f_W(y-x)dx\frac{1}{f_{Y^*}(y)}+\alpha_2 y^2+\lambda(y) \\ &= 0. \end{aligned} \tag{107}$$

Since the equations in (106) and (107) are satisfied for any x and y,

$$\begin{aligned} &-\log f_{Y^*}(y)+c_Y-\lambda(y)=0, \\ &\log f_{X^*}(x)+\alpha_0+\alpha_1 x^2+1-c_{Y^*}=0, \\ &\lambda(y)=1-\alpha_2 y^2, \end{aligned} \tag{108}$$

where c_Y is a constant.

Therefore,

$$\begin{aligned} f_{X^*}(x) &= \exp\left(-\alpha_0-\alpha_1 x^2-1+c_Y\right), \\ f_{Y^*}(y) &= \exp\left(c_Y-1+\alpha_2 y^2\right), \end{aligned}$$

and $f_{X^*}(x)$ and $f_{Y^*}(x)$ are re-written as

$$\begin{aligned} f_X(x) &= \exp\left(-\alpha_0-\alpha_1 x^2-1+c_Y\right) \\ &= \frac{1}{\sqrt{2\pi\frac{1}{2\alpha_1}}}\exp\left\{-\frac{1}{2\frac{1}{2\alpha_1}}x^2\right\}\sqrt{2\pi\frac{1}{2\alpha_1}}\exp\left\{-\alpha_0-1+c_Y\right\}, \end{aligned} \tag{109}$$

$$\begin{aligned} f_Y(y) &= \exp\left(c_Y-1+\alpha_2 y^2\right) \\ &= \frac{1}{\sqrt{2\pi\left(-\frac{1}{2\alpha_2}\right)}}\exp\left\{-\frac{1}{2\left(-\frac{1}{2\alpha_2}\right)}y^2\right\}\sqrt{2\pi\left(-\frac{1}{2\alpha_2}\right)}\exp\left\{c_Y-1\right\}. \end{aligned} \tag{110}$$

Considering the constraints in (103), the Lagrange multipliers in (109) and (110) are expressed as

$$\begin{aligned}
\alpha_0 &= -1 + c_{\scriptscriptstyle Y} + \frac{1}{2}\log 2\pi m_{\scriptscriptstyle X}^2 \\
&= \frac{1}{2}\log\frac{m_{\scriptscriptstyle X}^2}{m_{\scriptscriptstyle Y}^2}, \\
\alpha_1 &= \frac{1}{2m_{\scriptscriptstyle X}^2}, \\
\alpha_2 &= -\frac{1}{2m_{\scriptscriptstyle Y}^2}, \\
c_{\scriptscriptstyle Y} &= 1 - \frac{1}{2}\log 2\pi m_{\scriptscriptstyle Y}^2. \qquad (111)
\end{aligned}$$

Therefore, Gaussian density functions f_{X^*} and f_{Y^*} satisfy the first-order variation condition, $\delta U = 0$.

Now, the second-order variation condition must be considered, and, for the minimum, it requires the positive definiteness of the matrix,

$$\left.\begin{bmatrix} K''_{f_X f_X} & K''_{f_X f_Y} \\ K''_{f_Y f_X} & K''_{f_Y f_Y} \end{bmatrix}\right|_{f_X=f_{X^*}, f_Y=f_{Y^*}}. \qquad (112)$$

The elements of the matrix in (112) are calculated as

$$\begin{aligned}
\left.K''_{f_X f_X}\right|_{f_X=f_{X^*}, f_Y=f_{Y^*}} &= \frac{f_W(y-x)}{f_{X^*}(x)}, \\
\left.K''_{f_Y f_X}\right|_{f_X=f_{X^*}, f_Y=f_{Y^*}} &= -\frac{f_W(y-x)}{f_{Y^*}(x)}, \\
\left.K''_{f_X f_Y}\right|_{f_X=f_{X^*}, f_Y=f_{Y^*}} &= -\frac{f_W(y-x)}{f_{Y^*}(x)}, \\
\left.K''_{f_Y f_Y}\right|_{f_X=f_{X^*}, f_Y=f_{Y^*}} &= \frac{f_W(y-x) f_{X^*}(x)}{f_{Y^*}(y)^2}, \qquad (113)
\end{aligned}$$

and the matrix in (112) is positive definite. Therefore, $\delta^2 U > 0$, the optimal solutions f_{X^*} and f_{Y^*} minimize the variational problem in (102). Even though the optimal solutions are necessarily optimal, there are only Gaussian density functions f_{X^*} and f_{Y^*} in the feasible set, i.e., Gaussian density functions f_{X^*} and f_{Y^*} are the only ones which satisfy the equations in (106) and (107). Therefore, these optimal solutions are actually sufficient.

In conclusion, given the second-order moment, a Gaussian random variable $X_{\scriptscriptstyle G}$ minimizes the mutual information $I(X + W_{\scriptscriptstyle G}; W_{\scriptscriptstyle G})$, and the proof is completed. ■

APPENDIX I
PROOF OF THEOREM 12

Proof: To prove the inequality in (35), we first construct a functional problem as follows:

$$\begin{aligned}\min_{f_X} \quad & -\int\int f_X(\mathbf{x}) f_{Y|X}(\mathbf{y}|\mathbf{x}) \log\left(\int f_X(\mathbf{x}) f_{Y|X}(\mathbf{y}|\mathbf{x}) d\mathbf{x}\right) d\mathbf{x}d\mathbf{y} \qquad (114)\\ & +\int\int f_X(\mathbf{x}) f_{Y|X}(\mathbf{y}|\mathbf{x}) \log f_X(\mathbf{x}) d\mathbf{x}d\mathbf{y}\\ \text{s. t.} \quad & \int f_X(\mathbf{x}) d\mathbf{x} = 1,\\ & \int \mathbf{x} f_X(\mathbf{x}) d\mathbf{x} = \boldsymbol{\mu}_X,\\ & \int \mathbf{x}\mathbf{x}^T f_X(\mathbf{x}) d\mathbf{x} = \boldsymbol{\Omega}_X. \qquad (115)\end{aligned}$$

By substituting the random vector $\mathbf{Y}$ for $\mathbf{X} + \mathbf{W}_G$, where $\mathbf{X}$ and $\mathbf{W}_G$ are independent of each other, in (35), its density function $f_Y(\mathbf{y})$ and conditional density function $f_{Y|X}(\mathbf{y}|\mathbf{x})$ are expressed as

$$f_Y(\mathbf{y}) = \int f_X(\mathbf{x}) f_{Y|X}(\mathbf{y}|\mathbf{x}) d\mathbf{x}, \qquad (116)$$

$$f_{Y|X}(\mathbf{y}|\mathbf{x}) = f_W(\mathbf{y} - \mathbf{x}), \qquad (117)$$

respectively. Therefore, by substituting $f_Y(\mathbf{y})$ for $\int f_X(\mathbf{x}) f_{Y|X}(\mathbf{y}|\mathbf{x}) d\mathbf{x}$ and $f_W(\mathbf{y}-\mathbf{x})$ for $f_{Y|X}(\mathbf{y}|\mathbf{x})$, and appropriately changing the constrains in (115), the variational problem in (114) can be expressed as

$$\begin{aligned}\min_{f_X, f_Y} \quad & \int\int f_X(\mathbf{x}) f_W(\mathbf{y}-\mathbf{x}) \left[-\log f_Y(\mathbf{y}) + \log f_X(\mathbf{x})\right] d\mathbf{x}d\mathbf{y} \qquad (118)\\ \text{s. t.} \quad & \int\int f_X(\mathbf{x}) f_W(\mathbf{y}-\mathbf{x}) d\mathbf{x}d\mathbf{y} = 1,\\ & \int\int \mathbf{x} f_X(\mathbf{x}) f_W(\mathbf{y}-\mathbf{x}) d\mathbf{x}d\mathbf{y} = \boldsymbol{\mu}_X,\\ & \int\int \mathbf{x}\mathbf{x}^T f_X(\mathbf{x}) f_W(\mathbf{y}-\mathbf{x}) d\mathbf{x}d\mathbf{y} = \boldsymbol{\Omega}_X,\\ & \int f_Y(\mathbf{y}) d\mathbf{y} = 1,\\ & \int \mathbf{y} f_Y(\mathbf{y}) d\mathbf{y} = \boldsymbol{\mu}_Y,\\ & \int \mathbf{y}\mathbf{y}^T f_Y(\mathbf{x}) d\mathbf{y} = \boldsymbol{\Omega}_Y,\\ & f_Y(\mathbf{y}) = \int f_X(\mathbf{x}) f_W(\mathbf{y}-\mathbf{x}) d\mathbf{x}. \qquad (119)\end{aligned}$$

The functional problem in (118) is changed into the following equivalent problem:

$$\min_{f_X, f_Y} \int \left(\int f_X(\mathbf{x}) f_W(\mathbf{y}-\mathbf{x}) \left[-\log f_Y(\mathbf{y}) + \log f_X(\mathbf{x}) + \alpha_0 + \sum_{i=1}^{n} \zeta_i x_i + \sum_{i=1}^{n}\sum_{j=1}^{n} \gamma_{ij} x_i x_j - \lambda(\mathbf{y}) \right] d\mathbf{x} \right)$$
$$+ f_Y(\mathbf{y}) \left[\alpha_1 + \sum_{i=1}^{n} \eta_i y_i + \sum_{i=1}^{n}\sum_{j=1}^{n} \theta_{ij} y_i y_j + \lambda(\mathbf{y}) \right] d\mathbf{y}, \tag{120}$$

where $\mathbf{x}^T = [x_1, \ldots, x_n]$, $\mathbf{y}^T = [y_1, \ldots, y_n]$, and α_0, α_1, ζ_i, γ_{ij}, η_i, θ_{ij}, and $\lambda(\mathbf{y})$ are Lagrange multipliers.

Let's define the functional U as

$$U[f_X, f_Y] \quad = \quad \int \left(\int K(\mathbf{x}, \mathbf{y}, f_X, f_Y) d\mathbf{x} \right) + \tilde{K}(\mathbf{y}, f_Y) d\mathbf{y},$$

where

$$\begin{aligned} K(\mathbf{x}, \mathbf{y}, f_X, f_Y) &= f_X(\mathbf{x}) f_W(\mathbf{y}-\mathbf{x}) [-\log f_Y(\mathbf{y}) + \log f_X(\mathbf{x}) + \alpha_0 + \sum_{i=1}^{n} \zeta_i x_i + \sum_{i=1}^{n}\sum_{j=1}^{n} \gamma_{ij} x_i x_j - \lambda(\mathbf{y})], \\ \tilde{K}(\mathbf{y}, f_Y) &= f_Y(\mathbf{y}) \left[\alpha_1 + \sum_{i=1}^{n} \eta_i y_i + \sum_{i=1}^{n}\sum_{j=1}^{n} \theta_{ij} y_i y_j + \lambda(\mathbf{y}) \right]. \end{aligned} \tag{121}$$

Based on the first-order variation condition, we can find the optimal solution, f_{X^*} and f_{Y^*}, as follows.

$$\begin{aligned} & K'_{f_X}\Big|_{f_X = f_{X^*}, f_Y = f_{Y^*}} \\ = \; & f_W(\mathbf{y}-\mathbf{x}) \Bigg(-\log f_{Y^*}(\mathbf{y}) + \log f_{X^*}(\mathbf{x}) + \alpha_0 + \sum_{i=1}^{n} \zeta_i x_i \\ & \qquad + \sum_{i=1}^{n}\sum_{j=1}^{n} \gamma_{ij} x_i x_j + 1 - \lambda(\mathbf{y}) \Bigg) \\ = \; & f_W(\mathbf{y}-\mathbf{x}) \left(-\log f_{Y^*}(\mathbf{y}) + \log f_{X^*}(\mathbf{x}) + \alpha_0 + \boldsymbol{\zeta}\mathbf{x}^T + \mathbf{x}^T \boldsymbol{\Gamma} \mathbf{x} + 1 - \lambda(\mathbf{y}) \right) \\ = \; & 0 \end{aligned} \tag{122}$$

$$\begin{aligned} & \int K'_{f_Y} d\mathbf{x} + \tilde{K}'_{f_Y} \Big|_{f_X = f_{X^*}, f_Y = f_{Y^*}} \\ = \; & -\int f_{X^*}(\mathbf{x}) f_W(\mathbf{y}-\mathbf{x}) d\mathbf{x} \frac{1}{f_{Y^*}(\mathbf{y})} + \alpha_1 + \sum_{i=1}^{n} \eta_i y_i + \sum_{i=1}^{n}\sum_{j=1}^{n} \theta_{ij} y_i y_j + \lambda(\mathbf{y}) \\ = \; & -\int f_{X^*}(\mathbf{x}) f_W(\mathbf{y}-\mathbf{x}) d\mathbf{x} \frac{1}{f_{Y^*}(\mathbf{y})} + \alpha_1 + \boldsymbol{\eta}^T \mathbf{y} + \mathbf{y}^T \boldsymbol{\Theta} \mathbf{y} + \lambda(\mathbf{y}) \\ = \; & 0, \end{aligned} \tag{123}$$

where

$$\mathbf{\Gamma} = \begin{bmatrix} \gamma_{11} & \cdots & \gamma_{1n} \\ \vdots & \ddots & \vdots \\ \gamma_{n1} & \cdots & \gamma_{nn} \end{bmatrix}, \quad \mathbf{\Theta} = \begin{bmatrix} \theta_{11} & \cdots & \theta_{1n} \\ \vdots & \ddots & \vdots \\ \theta_{n1} & \cdots & \theta_{nn} \end{bmatrix}, \tag{124}$$

$\boldsymbol{\zeta} = [\zeta_1, \ldots, \zeta_n]^T$, and $\boldsymbol{\eta} = [\eta_1, \ldots, \eta_n]^T$.

Since the equalities in (122) and (123) must be satisfied for any $\mathbf{x}$ and $\mathbf{y}$,

$$\begin{aligned} 0 &= -\log f_{Y^*}(\mathbf{y}) - \lambda(\mathbf{y}), \\ 0 &= \log f_{X^*}(\mathbf{x}) + \alpha_0 + \boldsymbol{\zeta}\mathbf{x}^T + \mathbf{x}^T\mathbf{\Gamma}\mathbf{x} + 1, \\ \lambda(\mathbf{y}) &= 1 - \alpha_1 - \boldsymbol{\eta}^T\mathbf{y} - \mathbf{y}^T\mathbf{\Theta}\mathbf{y}, \end{aligned} \tag{125}$$

and

$$\begin{aligned} f_{X^*}(\mathbf{x}) &= \exp\left(-\alpha_0 - \boldsymbol{\zeta}^T\mathbf{x} - \mathbf{x}^T\mathbf{\Gamma}\mathbf{x} - 1\right), \\ f_{Y^*}(\mathbf{y}) &= \exp\left(-1 + \alpha_1 + \boldsymbol{\eta}^T\mathbf{y} + \mathbf{y}^T\mathbf{\Theta}\mathbf{y}\right). \end{aligned} \tag{126}$$

Considering the constraints in (119), $f_{X^*}(\mathbf{x})$ and $f_{Y^*}(\mathbf{x})$ in (126) are expressed as

$$\begin{aligned} f_{X^*}(\mathbf{x}) &= (2\pi)^{-\frac{n}{2}} |\mathbf{\Sigma}_X|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_X)^T \mathbf{\Sigma}_X^{-1} (\mathbf{x} - \boldsymbol{\mu}_X)\right\}, \\ &= \exp\left\{-\frac{1}{2}\log(2\pi)^n |\mathbf{\Sigma}_X| - \frac{1}{2}\mathbf{x}^T\mathbf{\Sigma}_X^{-1}\mathbf{x} + \boldsymbol{\mu}_X^T\mathbf{\Sigma}_X^{-1}\mathbf{x} - \frac{1}{2}\boldsymbol{\mu}_X^T\mathbf{\Sigma}_X^{-1}\boldsymbol{\mu}_X\right\} \\ &= \exp\left(-\alpha_0 - \boldsymbol{\zeta}^T\mathbf{x} - \mathbf{x}^T\mathbf{\Gamma}\mathbf{x} - 1\right), \\ f_{Y^*}(\mathbf{y}) &= (2\pi)^{-\frac{n}{2}} |\mathbf{\Sigma}_Y|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\mathbf{y} - \boldsymbol{\mu}_Y)^T \mathbf{\Sigma}_Y^{-1} (\mathbf{y} - \boldsymbol{\mu}_Y)\right\} \\ &= \exp\left\{-\frac{1}{2}\log(2\pi)^n |\mathbf{\Sigma}_Y| - \frac{1}{2}\mathbf{y}^T\mathbf{\Sigma}_Y^{-1}\mathbf{y} + \boldsymbol{\mu}_Y^T\mathbf{\Sigma}_Y^{-1}\mathbf{y} - \frac{1}{2}\boldsymbol{\mu}_Y^T\mathbf{\Sigma}_Y^{-1}\boldsymbol{\mu}_Y\right\} \\ &= \exp\left(-1 + \alpha_1 + \boldsymbol{\eta}^T\mathbf{y} + \mathbf{y}^T\mathbf{\Theta}\mathbf{y}\right), \end{aligned} \tag{127}$$

where $\mathbf{\Sigma}_X = \mathbf{\Omega}_X - \boldsymbol{\mu}_X\boldsymbol{\mu}_X^T$, $\mathbf{\Sigma}_Y = \mathbf{\Sigma}_X + \mathbf{\Sigma}_W$, and $\mathbf{\Sigma}_W$ is a covariance matrix of $\mathbf{W}_G$. Based on the

equations in (127),

$$\begin{aligned}
\alpha_0 &= -1 + \frac{1}{2}\log{(2\pi)^n}\left|\boldsymbol{\Sigma}_X\right| + \frac{1}{2}\boldsymbol{\mu}_X^T\boldsymbol{\Sigma}_X^{-1}\boldsymbol{\mu}_X, \\
\alpha_1 &= 1 - \frac{1}{2}\log{(2\pi)^n}\left|\boldsymbol{\Sigma}_Y\right| - \frac{1}{2}\boldsymbol{\mu}_Y^T\boldsymbol{\Sigma}_Y^{-1}\boldsymbol{\mu}_Y, \\
\boldsymbol{\Gamma} &= \frac{1}{2}\boldsymbol{\Sigma}_X^{-1}, \\
\boldsymbol{\zeta} &= -\boldsymbol{\mu}_X^T\boldsymbol{\Sigma}_X^{-1}, \\
\boldsymbol{\Theta} &= \frac{1}{2}\boldsymbol{\Sigma}_Y^{-1}, \\
\boldsymbol{\eta} &= -\boldsymbol{\mu}_Y^T\boldsymbol{\Sigma}_Y^{-1}.
\end{aligned} \tag{128}$$

Therefore, the optimal solutions f_{X^*} and f_{Y^*} are multi-variate Gaussian density functions (without loss of generality, we assume that the covariance matrix $\boldsymbol{\Sigma}_X$ is invertible due to the reason mentioned in Appendix B).

Now, by confirming the second-order variation condition, we will show that the optimal solutions f_{X^*} and f_{Y^*} minimize the variational functional in (118). Based on Theorem 2, we will show that the following matrix is positive definite:

$$\begin{bmatrix} K''_{f_X f_X} & K''_{f_X f_Y} \\ K''_{f_Y f_X} & K''_{f_Y f_Y} \end{bmatrix} > \mathbf{0}. \tag{129}$$

Since the elements of the matrix in (129) are defined as

$$\begin{aligned}
\left.K''_{f_X f_X}\right|_{f_X=f_{X^*}, f_Y=f_{Y^*}} &= \frac{f_W(\mathbf{y}-\mathbf{x})}{f_{X^*}(\mathbf{x})}, \\
\left.K''_{f_Y f_Y}\right|_{f_X=f_{X^*}, f_Y=f_{Y^*}} &= \frac{f_{X^*}(\mathbf{x})f_W(\mathbf{y}-\mathbf{x})}{f_{Y^*}(\mathbf{y})^2}, \\
\left.K''_{f_X f_Y}\right|_{f_X=f_{X^*}, f_Y=f_{Y^*}} &= -\frac{f_W(\mathbf{y}-\mathbf{x})}{f_{Y^*}(\mathbf{y})}, \\
\left.K''_{f_Y f_X}\right|_{f_X=f_{X^*}, f_Y=f_{Y^*}} &= -\frac{f_W(\mathbf{y}-\mathbf{x})}{f_{Y^*}(\mathbf{y})},
\end{aligned} \tag{130}$$

the matrix is a positive definite matrix, and therefore $\delta^2 U > 0$. Therefore, the optimal solutions f_{X^*} and f_{Y^*} actually minimize the variational functional in (118). Even though these optimal solutions are necessarily optimal, there exists only one solution, which is a multi-variate Gaussian density function, which satisfies Euler's equation in (122) and (123). Therefore, f_{X^*} and f_{Y^*} are also sufficient in this case.

Remark 11: The constraints related to the mean vectors in (119) are unnecessary. Without these constraints, the optimal solutions are still multi-variate Gaussian density functions but the mean vectors are changed into zero.

■

APPENDIX J

PROOF OF THEOREM 13

Proof: To prove the entropy power inequality, we slightly change the inequality in (36) into the following relationship:

$$h(\tilde{X}+\tilde{W}) \geq a_X^2 h(\tilde{X}) + a_W^2 h(\tilde{W}) - \log a_X - \log a_W, \tag{131}$$

where $\tilde{X} = a_X X$ and $\tilde{X} = a_W W$. Since a_X and a_W are constants, they do not affect the optimization, and we can ignore these two terms.

Based on the inequality in (131) and required constraints, construct the following functional problem (for the simplicity of the notation, we simply denote $\tilde{X}$ and $\tilde{W}$ as X and W):

$$\min_{f_X, f_W, f_Y} \quad \int\int f_X(x) f_W(y-x) \left(-\log f_Y(y) + a_X^2 \log f_X(x) + a_W^2 \log f_W(y-x)\right) dxdy \tag{132}$$

$$\begin{aligned}
\text{s.t.} \quad & \int\int f_X(x) f_W(y-x)\, dxdy = 1, \\
& \int\int y^2 f_X(x) f_W(y-x)\, dxdy = m_{Y^*}^2, \\
& \int\int x^2 f_X(x) f_W(y-x) dxdy = m_{X^*}^2, \\
& \int\int (y-x)^2 f_X(x) f_W(y-x) dxdy = m_{W^*}^2, \\
& -\int\int f_X(x) f_W(y-x) \log f_X(x) dxdy = p_X, \\
& -\int\int f_X(x) f_W(y-x) \log f_W(y-x) dxdy = p_W, \\
& f_Y(y) = \int f_X(x) f_W(y-x)\, dx,
\end{aligned} \tag{133}$$

where $m_{X^*}^2$, $m_{W^*}^2$, and $m_{Y^*}^2$ denote the second-order moments of the optimal solutions of X, W, and Y, respectively. The constraints related to the second-order moments mean that all random variables have finite second-order moments. Also, the constraints related to p_X and p_W mean that random variables X and W have finite entropies, respectively, where p_X and p_W are constants. Without loss of generality, the zero mean condition is assumed for all random variables (in the case of non-zero mean, all constraints related to the second-order moments are changed into constraints related to the covariance matrices).

Using Lagrange multipliers, the problem in (132) and the constraints in (133) are reformulated as the following equivalent problem:

$$\min_{f_X, f_W, f_Y} \int \left(\int K(x, y, f_X, f_W, f_Y) dx \right) + \tilde{K}(y, f_Y) dy, \tag{134}$$

where

$$\begin{aligned} K(x, y, f_X, f_W, f_Y) &= f_X(x) f_W(y-x) \Big(-\log f_Y(y) + (a_X^2 - \lambda_X) \log f_X(x) \\ &\quad + (a_W^2 - \lambda_W) \log f_W(y-x) + \alpha_0 + \alpha_1 y^2 + \alpha_2 x^2 + \alpha_3 (y-x)^2 - \lambda(y) \Big), \\ \tilde{K}(y, f_Y) &= \lambda(y) f_Y(y). \end{aligned} \tag{135}$$

The first-order partial derivative is expressed as

$$\begin{aligned} & K'_{f_X}\Big|_{f_X = f_{X^*}, f_W = f_{W^*}, f_Y = f_{Y^*}} \\ = \; & f_{W^*}(y-x) \Big(-\log f_{Y^*}(y) + (a_X^2 - \lambda_X) \log f_{X^*}(x) + (a_W^2 - \lambda_W) \log f_{W^*}(y-x) + \alpha_0 + \alpha_1 y^2 \\ & + \alpha_2 x^2 + \alpha_3 (y-x)^2 - \lambda(y) + a_X^2 - \lambda_X \Big), \\ & K'_{f_W}\Big|_{f_X = f_{X^*}, f_W = f_{W^*}, f_Y = f_{Y^*}} \\ = \; & f_{X^*}(x) \Big(-\log f_{Y^*}(y) + (a_X^2 - \lambda_X) \log f_{X^*}(x) + (a_W^2 - \lambda_W) \log f_{W^*}(y-x) + \alpha_0 + \alpha_1 y^2 \\ & + \alpha_2 x^2 + \alpha_3 (y-x)^2 - \lambda(y) + a_W^2 - \lambda_W \Big), \\ & \left(\int K dx + \tilde{K} \right)'_{f_Y} \Big|_{f_X = f_{X^*}, f_W = f_{W^*}, f_Y = f_{Y^*}} \\ = \; & -\int f_{X^*}(x) f_{W^*}(y-x) \, dx \frac{1}{f_{Y^*}(y)} + \lambda(y). \end{aligned} \tag{136}$$

Due to the first-order variation condition, $\delta U[f_{X^*}, f_{W^*}, f_{Y^*}] = 0$, the optimal solutions f_{X^*}, f_{W^*}, and f_{Y^*}, must satisfy the following relationships:

$$\begin{aligned} -\log f_{Y^*}(y) + \alpha_1 y^2 - \lambda(y) + c_Y &= 0, \\ (a_X^2 - \lambda_X) \log f_{X^*}(x) + \alpha_2 x^2 + c_X &= 0, \\ (a_W^2 - \lambda_W) \log f_{W^*}(y-x) + \alpha_3 (y-x)^2 + \alpha_0 + a_W^2 - \lambda_W - c_X - c_Y &= 0, \\ -1 + \lambda(y) &= 0, \\ a_W^2 - \lambda_W - a_X^2 + \lambda_X &= 0, \end{aligned} \tag{137}$$

and therefore,

$$
\begin{aligned}
f_{Y^*}(y) &= \exp\left\{\alpha_1 y^2 - \lambda(y) + c_Y\right\}, \\
f_{X^*}(x) &= \exp\left\{\frac{1}{a_X^2 - \lambda_X}\left(-\alpha_2 x^2 - c_X\right)\right\}, \\
f_{W^*}(y-x) &= \exp\left\{\frac{1}{a_W^2 - \lambda_W}\left(-\alpha_3 (y-x)^2 - \alpha_0 - a_W^2 + \lambda_W + c_X + c_Y\right)\right\}, \\
\lambda(y) &= 1.
\end{aligned} \tag{138}
$$

Considering the constraints in (133), the equations in (138) are expressed as

$$
\begin{aligned}
f_{Y^*}(y) &= \frac{1}{\sqrt{2\pi\left(-\frac{1}{2\alpha_1}\right)}}\exp\left\{-\frac{1}{2\left(-\frac{1}{2\alpha_1}\right)}y^2\right\}\sqrt{2\pi\left(-\frac{1}{2\alpha_1}\right)}\exp\left\{-\lambda(y) + c_Y\right\} \\
&= \frac{1}{\sqrt{2\pi m_{Y^*}^2}}\exp\left\{-\frac{1}{2m_{Y^*}^2}y^2\right\}, \\
f_{X^*}(x) &= \frac{1}{\sqrt{2\pi\left(\frac{a_X^2-\lambda_X}{2\alpha_2}\right)}}\exp\left\{-\frac{1}{2\left(\frac{a_X^2-\lambda_X}{2\alpha_2}\right)}x^2\right\}\sqrt{2\pi\left(\frac{a_X^2-\lambda_X}{2\alpha_2}\right)}\exp\left\{-\frac{c_X}{a_X^2-\lambda_X}\right\} \\
&= \frac{1}{\sqrt{2\pi m_{X^*}^2}}\exp\left\{-\frac{1}{2m_{X^*}^2}x^2\right\}, \\
f_{W^*}(y-x) &= \frac{1}{\sqrt{2\pi\left(\frac{a_W^2-\lambda_W}{2\alpha_3}\right)}}\exp\left\{-\frac{1}{2\left(\frac{a_W^2-\lambda_W}{2\alpha_3}\right)}(y-x)^2\right\} \\
&\quad \times\sqrt{2\pi\left(\frac{a_W^2-\lambda_W}{2\alpha_3}\right)}\exp\left\{\frac{-\alpha_0 - a_W^2 + \lambda_W + c_X + c_Y}{a_W^2-\lambda_W}\right\} \\
&= \frac{1}{\sqrt{2\pi m_{W^*}^2}}\exp\left\{-\frac{1}{2m_{W^*}^2}(y-x)^2\right\},
\end{aligned} \tag{139}
$$

where

$$
\begin{aligned}
\alpha_0 &= -(a_W^2-\lambda_W)+c_X+c_Y+\frac{a_W^2-\lambda_W}{2}\log 2\pi m_{W^*}^2 \\
\alpha_1 &= -\frac{1}{2m_{Y^*}^2}, \\
\alpha_2 &= \frac{a_X^2-\lambda_X}{2m_{X^*}^2}, && (140)\\
\alpha_3 &= \frac{a_W^2-\lambda_W}{2m_{W^*}^2}, && (141)\\
c_X &= \frac{a_X^2-\lambda_X}{2}\log 2\pi m_{X^*}^2 \\
c_Y &= 1-\frac{1}{2}\log 2\pi m_{Y^*}^2, \\
a_X^2-\lambda_X &= a_W^2-\lambda_W \geq 1, && (142)\\
m_{X^*}^2 &= \frac{1}{2\pi e}\exp\left\{2p_X\right\}, \\
m_{W^*}^2 &= \frac{1}{2\pi e}\exp\left\{2p_W\right\}, \\
m_{Y^*}^2 &= m_{X^*}^2+m_{W^*}^2 \\
&= \frac{1}{2\pi e}\exp\left\{2p_X\right\}+\frac{1}{2\pi e}\exp\left\{2p_W\right\}.
\end{aligned}
$$

The inequality in (142) is due to the second-order variation condition, which will be justified next.

Consider now the conditions for the second variation of the functional problem:

$$
\begin{aligned}
K''_{f_X f_X}\Big|_{f_X=f_{X^*},f_W=f_{W^*},f_Y=f_{Y^*}} &= \frac{(a_X^2-\lambda_X)f_{W^*}(y-x)}{f_{X^*}(x)}, \\
K''_{f_W f_W}\Big|_{f_X=f_{X^*},f_W=f_{W^*},f_Y=f_{Y^*}} &= \frac{(a_W^2-\lambda_W)f_{X^*}(x)}{f_{W^*}(y-x)}, \\
\left(\int K dx+\tilde{K}\right)''_{f_Y f_Y}\Big|_{f_X=f_{X^*},f_W=f_{W^*},f_Y=f_{Y^*}} &= \frac{f_{X^*}(x)f_{W^*}(y-x)}{f_{Y^*}(y)^2}, \\
K''_{f_X f_W}\Big|_{f_X=f_{X^*},f_W=f_{W^*},f_Y=f_{Y^*}} &= a_W^2-\lambda_W, \\
K''_{f_W f_X}\Big|_{f_X=f_{X^*},f_W=f_{W^*},f_Y=f_{Y^*}} &= a_X^2-\lambda_X, \\
K''_{f_X f_Y}\Big|_{f_X=f_{X^*},f_W=f_{W^*},f_Y=f_{Y^*}} &= -\frac{f_{W^*}(y-x)}{f_{Y^*}(y)}, \\
K''_{f_Y f_X}\Big|_{f_X=f_{X^*},f_W=f_{W^*},f_Y=f_{Y^*}} &= -\frac{f_{W^*}(y-x)}{f_{Y^*}(y)}, \\
K''_{f_W f_Y}\Big|_{f_X=f_{X^*},f_W=f_{W^*},f_Y=f_{Y^*}} &= -\frac{f_{X^*}(x)}{f_{Y^*}(y)}, \\
K''_{f_Y f_W}\Big|_{f_X=f_{X^*},f_W=f_{W^*},f_Y=f_{Y^*}} &= -\frac{f_{X^*}(x)}{f_{Y^*}(y)}. && (143)
\end{aligned}
$$

To satisfy $\delta^2 J \geq 0$, the following condition must hold:

$$
\begin{aligned}
&\left[\begin{array}{ccc} h_X & h_W & h_Y \end{array}\right]\left[\begin{array}{ccc} K''_{f_X f_X} & K''_{f_X f_W} & K''_{f_X f_Y} \\ K''_{f_W f_X} & K''_{f_W f_W} & K''_{f_W f_Y} \\ K''_{f_Y f_X} & K''_{f_Y f_W} & K''_{f_Y f_Y} \end{array}\right]\left[\begin{array}{c} h_X \\ h_W \\ h_Y \end{array}\right]\Bigg|_{f_X=f_{X^*}, f_W=f_{W^*}, f_Y=f_{Y^*}} \\
= \;& K''_{f_X f_X} h_X^2 + K''_{f_W f_W} h_W^2 + K''_{f_Y f_Y} h_Y^2 + (K''_{f_X f_W} + K''_{f_W f_X}) h_X h_W \\
& + (K''_{f_W f_Y} + K''_{f_Y f_W}) h_W h_Y + (K''_{f_X f_Y} + K''_{f_Y f_X}) h_Y h_X \Big|_{f_X=f_{X^*}, f_W=f_{W^*}, f_Y=f_{Y^*}} \\
\geq \;& 0.
\end{aligned} \tag{144}
$$

Using the defined quantities in (143), the equation in (144) is expressed as follows:

$$
\begin{aligned}
& K''_{f_{X^*} f_{X^*}} h_X^2 + K''_{f_{W^*} f_{W^*}} h_W^2 + K''_{f_{Y^*} f_{Y^*}} h_Y^2 \\
& + (K''_{f_{X^*} f_{W^*}} + K''_{f_{W^*} f_{X^*}}) h_X h_W + (K''_{f_{W^*} f_{Y^*}} + K''_{f_{Y^*} f_{W^*}}) h_W h_Y + (K''_{f_{X^*} f_{Y^*}} + K''_{f_{Y^*} f_{X^*}}) h_Y h_X \\
= \;& \frac{(a_X^2 - \lambda_X) f_{W^*}(y-x)}{f_{X^*}(x)} h_X(x)^2 + \frac{(a_W^2 - \lambda_W) f_{X^*}(x)}{f_{W^*}(y-x)} h_W(y-x)^2 + \frac{f_{X^*}(x) f_{W^*}(y-x)}{f_{Y^*}(y)^2} h_Y(y)^2 \\
& + 2(a_W^2 - \lambda_W) h_X(x) h_W(y-x) - 2\frac{f_{X^*}(x)}{f_{Y^*}(y)} h_W(y-x) h_Y(y) - 2\frac{f_{W^*}(y-x)}{f_{Y^*}(y)} h_X(x) h_Y(y) \\
= \;& \frac{f_{W^*}(y-x)}{f_{X^*}(x)} \Bigg((a_W^2 - \lambda_W) h_X(x)^2 + (a_W^2 - \lambda_W) \frac{f_{X^*}(x)^2}{f_{W^*}(y-x)^2} h_W(y-x)^2 + \frac{f_{X^*}(x)^2}{f_{Y^*}(y)^2} h_Y(y)^2 \\
& \qquad + 2(a_W^2 - \lambda_W) \frac{f_{X^*}(x)}{f_{W^*}(y-x)} h_X(x) h_W(y-x) \\
& \qquad - 2\frac{f_{X^*}(x)^2}{f_{W^*}(y-x) f_{Y^*}(y)} h_W(y-x) h_Y(y) - 2\frac{f_{X^*}(x)}{f_{Y^*}(y)} h_X(x) h_Y(y) \Bigg) \\
= \;& \frac{f_{W^*}(y-x)}{f_{X^*}(x)} \left(h_X(x) + \frac{f_{X^*}(x)}{f_{W^*}(y-x)} h_W(y-x) - \frac{f_{X^*}(x)}{f_{Y^*}(y)} h_Y(y) \right)^2 \\
\geq \;& 0,
\end{aligned} \tag{145}
$$

where $a_W^2 - \lambda_W = a_X^2 - \lambda_X \geq 1$.

Therefore, the optimal solutions, f_{X^*}, f_{W^*}, and f_{Y^*}, minimize the variational problem in (132). Even though f_{X^*}, f_{W^*}, and f_{Y^*} are necessarily optimal, they are sufficiently optimal since only Gaussian density functions are in the feasible constraints set. ■

APPENDIX K

PROOF OF THEOREM 14

Proof: Similar to the proof shown in Appendix J, we first construct the following functional problem:

$$\begin{aligned}
\min_{f_X, f_W, f_Y} \quad & \int\int f_X(\mathbf{x}) f_W(\mathbf{y}-\mathbf{x}) \left(-\log f_Y(\mathbf{y}) + a_X^2 \log f_X(\mathbf{x}) + a_W^2 \log f_W(\mathbf{y}-\mathbf{x})\right) d\mathbf{x} d\mathbf{y} \\
\text{s.t.} \quad & \int\int f_X(\mathbf{x}) f_W(\mathbf{y}-\mathbf{x})\, d\mathbf{x} d\mathbf{y} = 1, \\
& \int\int \mathbf{y}\mathbf{y}^T f_X(\mathbf{x}) f_W(\mathbf{y}-\mathbf{x})\, d\mathbf{x} d\mathbf{y} = \mathbf{\Omega}_{X^*} + \mathbf{\Omega}_{W^*}, \\
& \int\int \mathbf{x}\mathbf{x}^T f_X(\mathbf{x}) f_W(\mathbf{y}-\mathbf{x})\, d\mathbf{x} d\mathbf{y} = \mathbf{\Omega}_{X^*}, \\
& \int\int (\mathbf{y}-\mathbf{x})(\mathbf{y}-\mathbf{x})^T f_X(\mathbf{x}) f_W(\mathbf{y}-\mathbf{x})\, d\mathbf{x} d\mathbf{y} = \mathbf{\Omega}_{W^*}, \\
& -\int\int f_X(\mathbf{x}) f_W(\mathbf{y}-\mathbf{x}) \log f_X(\mathbf{x}) d\mathbf{x} d\mathbf{y} = p_X, \\
& -\int\int f_X(\mathbf{x}) f_W(\mathbf{y}-\mathbf{x}) \log f_W(\mathbf{y}-\mathbf{x})\, d\mathbf{x} d\mathbf{y} = p_W, \\
& f_Y(\mathbf{y}) = \int f_X(\mathbf{x}) f_W(\mathbf{y}-\mathbf{x})\, d\mathbf{x},
\end{aligned} \tag{146}$$

where p_X and p_W are constants, and the constraints related to these constants mean the entropies of $\mathbf{X}$ and $\mathbf{W}$ are finite. The matrices $\mathbf{\Omega}_{X^*}$ and $\mathbf{\Omega}_{W^*}$ denote the correlation matrices of the optimal random vectors $\mathbf{X}^*$ and $\mathbf{W}^*$, respectively. The constraints related to these correlation matrices mean that the correlation matrices of random vectors $\mathbf{X}$ and $\mathbf{W}$ exist. Without loss of generality, the mean vectors of $\mathbf{X}$ and $\mathbf{W}$ are assumed to be zero (If $\mathbf{X}$ and $\mathbf{W}$ have non-zero mean vectors, the constraints related to the correlation matrices are changed into the ones related to the covariance matrices.).

Using Lagrange multipliers, the problem in (146) is changed into the following optimization problem:

$$\min_{f_X, f_W, f_Y} \quad \int \left(\int K(\mathbf{x}, \mathbf{y}, f_X, f_W, f_Y) d\mathbf{x} \right) + \tilde{K}(\mathbf{y}, f_Y) d\mathbf{y}, \tag{147}$$

where

$$\begin{aligned}
K(\mathbf{x}, \mathbf{y}, f_X, f_W, f_Y) &= f_X(\mathbf{x}) f_W(\mathbf{y}-\mathbf{x}) \Big(-\log f_Y(\mathbf{y}) + (a_X^2 - \lambda_X) \log f_X(\mathbf{x}) \\
&\quad + (a_W^2 - \lambda_W) \log f_W(\mathbf{y}-\mathbf{x}) + \alpha + \sum_{i=1}^{n}\sum_{j=1}^{n} \gamma_{ij} y_i y_j \\
&\quad + \sum_{i=1}^{n}\sum_{j=1}^{n} \phi_{ij} x_i x_j + \sum_{i=1}^{n}\sum_{j=1}^{n} \theta_{ij} (y_i - x_i)(y_j - x_j) - \lambda(\mathbf{y}) \Big), \\
\tilde{K}(\mathbf{y}, f_Y) &= \lambda(\mathbf{y}) f_Y(\mathbf{y}).
\end{aligned} \tag{148}$$

Then,

$$
\begin{aligned}
K'_{f_X} &= f_W\left(\mathbf{y}-\mathbf{x}\right)\Big(-\log f_Y(\mathbf{y}) + (a_X^2-\lambda_X)\log f_X(\mathbf{x}) + (a_W^2-\lambda_W)\log f_W\left(\mathbf{y}-\mathbf{x}\right) + \alpha \\
&\quad + \sum_{i=1}^{n}\sum_{j=1}^{n}\gamma_{ij}y_iy_j + \sum_{i=1}^{n}\sum_{j=1}^{n}\phi_{ij}x_ix_j + \sum_{i=1}^{n}\sum_{j=1}^{n}\theta_{ij}\left(y_i-x_i\right)\left(y_j-x_j\right) - \lambda(\mathbf{y}) + a_X^2 - \lambda_X\Big), \\
K'_{f_W} &= f_X(\mathbf{x})\Big(-\log f_Y(\mathbf{y}) + (a_X^2-\lambda_X)\log f_X(\mathbf{x}) + (a_W^2-\lambda_W)\log f_W\left(\mathbf{y}-\mathbf{x}\right) + \alpha \\
&\quad + \sum_{i=1}^{n}\sum_{j=1}^{n}\gamma_{ij}y_iy_j + \sum_{i=1}^{n}\sum_{j=1}^{n}\phi_{ij}x_ix_j + \sum_{i=1}^{n}\sum_{j=1}^{n}\theta_{ij}\left(y_i-x_i\right)\left(y_j-x_j\right) - \lambda(\mathbf{y}) + a_W^2 - \lambda_W\Big), \\
\left(\int K d\mathbf{x} + \tilde{K}\right)'_{f_Y} &= -\int f_X(\mathbf{x})f_W\left(\mathbf{y}-\mathbf{x}\right)d\mathbf{x}\frac{1}{f_Y(\mathbf{y})} + \lambda(\mathbf{y}).
\end{aligned}
\tag{149}
$$

To satisfy $\delta U[f_{X^*}, f_{W^*}, f_{Y^*}] = 0$,

$$
\begin{aligned}
& -\log f_{Y^*}(\mathbf{y}) + \sum_{i=1}^{n}\sum_{j=1}^{n}\gamma_{ij}y_iy_j - \lambda(\mathbf{y}) + c_Y = 0, \\
& (a_X^2-\lambda_X)\log f_{X^*}(\mathbf{x}) + \sum_{i=1}^{n}\sum_{j=1}^{n}\phi_{ij}x_ix_j + c_X = 0, \\
& (a_W^2-\lambda_W)\log f_{W^*}\left(\mathbf{y}-\mathbf{x}\right) + \sum_{i=1}^{n}\sum_{j=1}^{n}\theta_{ij}\left(y_i-x_i\right)\left(y_j-x_j\right) + \alpha + a_W^2 - \lambda_W - c_X - c_Y = 0, \\
& -1 + \lambda(\mathbf{y}) = 0, \\
& a_W^2 - \lambda_W - a_X^2 + \lambda_X = 0.
\end{aligned}
\tag{150}
$$

Since the equations in (150) must be satisfied for any $\mathbf{x}$ and $\mathbf{y}$, the optimal solutions f_{X^*}, f_{W^*}, and f_{W^*} are expressed as

$$
\begin{aligned}
f_{Y^*}(\mathbf{y}) &= \exp\left\{\sum_{i=1}^{n}\sum_{j=1}^{n}\gamma_{ij}y_iy_j - \lambda(\mathbf{y}) + c_Y\right\} \\
&= \exp\left\{\mathbf{y}^T\boldsymbol{\Gamma}\mathbf{y} - 1 + c_Y\right\}, \\
f_{X^*}(\mathbf{x}) &= \exp\left\{\frac{1}{a_X^2-\lambda_X}\left(-\sum_{i=1}^{n}\sum_{j=1}^{n}\phi_{ij}x_ix_j - c_X\right)\right\} \\
&= \exp\left\{-\frac{1}{a_X^2-\lambda_X}\left(\mathbf{x}^T\boldsymbol{\Phi}\mathbf{x} + c_X\right)\right\}, \\
f_{W^*}\left(\mathbf{y}-\mathbf{x}\right) &= \exp\left\{\frac{1}{a_W^2-\lambda_W}\left(-\sum_{i=1}^{n}\sum_{j=1}^{n}\theta_{ij}\left(y_i-x_i\right)\left(y_j-x_j\right) - \alpha - a_W^2 + \lambda_W + c_X + c_Y\right)\right\} \\
&= \exp\left\{-\frac{1}{a_W^2-\lambda_W}\left(\left(\mathbf{y}-\mathbf{x}\right)^T\boldsymbol{\Theta}\left(\mathbf{y}-\mathbf{x}\right) + \alpha + a_W^2 - \lambda_W - c_X - c_Y\right)\right\} \\
\lambda(\mathbf{y}) &= 1.
\end{aligned}
\tag{151}
$$

Considering the constraints in (146), the equations in (151) are further processed as

$$
\begin{aligned}
f_{Y^*}(\mathbf{y}) &= \frac{1}{(2\pi)^{\frac{n}{2}}\left|-\frac{1}{2}\mathbf{\Gamma}^{-1}\right|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}\mathbf{y}^T\left(-\frac{1}{2}\mathbf{\Gamma}^{-1}\right)^{-1}\mathbf{y}\right\}(2\pi)^{\frac{n}{2}}\left|-\frac{1}{2}\mathbf{\Gamma}^{-1}\right|^{\frac{1}{2}}\exp\left\{-\lambda(\mathbf{y})+c_Y\right\}\\
&= \frac{1}{(2\pi)^{\frac{n}{2}}\left|\mathbf{\Omega}_{X^*}+\mathbf{\Omega}_{W^*}\right|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}\mathbf{y}^T\left(\mathbf{\Omega}_{X^*}+\mathbf{\Omega}_{W^*}\right)^{-1}\mathbf{y}\right\},\\
f_{X^*}(\mathbf{x}) &= \frac{1}{(2\pi)^{\frac{n}{2}}\left|\frac{a_X^2-\lambda_X}{2}\mathbf{\Phi}^{-1}\right|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}\mathbf{x}^T\left(\frac{a_X^2-\lambda_X}{2}\mathbf{\Phi}^{-1}\right)^{-1}\mathbf{x}\right\}\\
&\quad\times(2\pi)^{\frac{n}{2}}\left|\frac{a_X^2-\lambda_X}{2}\mathbf{\Phi}^{-1}\right|^{\frac{1}{2}}\exp\left\{-\frac{c_X}{a_X^2-\lambda_X}\right\}\\
&= \frac{1}{(2\pi)^{\frac{n}{2}}\left|\mathbf{\Omega}_{X^*}\right|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}\mathbf{x}^T\mathbf{\Omega}_{X^*}^{-1}\mathbf{x}\right\},\\
f_{W^*}(\mathbf{y}-\mathbf{x}) &= \frac{1}{(2\pi)^{\frac{n}{2}}\left|\frac{a_W^2-\lambda_W}{2}\mathbf{\Theta}^{-1}\right|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}(\mathbf{y}-\mathbf{x})^T\left(\frac{a_W^2-\lambda_W}{2}\mathbf{\Theta}^{-1}\right)^{-1}(\mathbf{y}-\mathbf{x})\right\}\\
&\quad\times(2\pi)^{\frac{n}{2}}\left|\frac{a_W^2-\lambda_W}{2}\mathbf{\Theta}^{-1}\right|^{\frac{1}{2}}\exp\left\{\frac{-\alpha-a_W^2+\lambda_W+c_X+c_Y}{a_W^2-\lambda_W}\right\}\\
&= \frac{1}{(2\pi)^{\frac{n}{2}}\left|\mathbf{\Omega}_{W^*}\right|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}(\mathbf{y}-\mathbf{x})^T\mathbf{\Omega}_{W^*}^{-1}(\mathbf{y}-\mathbf{x})\right\},
\end{aligned}
\tag{152}
$$

where

$$
\begin{aligned}
\alpha &= -(a_W^2-\lambda_W)+c_X+c_Y+\frac{a_W^2-\lambda_W}{2}\log\left((2\pi)^n\left|\mathbf{\Omega}_{W^*}\right|\right)\\
\mathbf{\Gamma} &= -\frac{1}{2}\left(\mathbf{\Omega}_{X^*}+\mathbf{\Omega}_{W^*}\right)^{-1},\\
\mathbf{\Phi} &= \frac{a_X^2-\lambda_X}{2}\mathbf{\Omega}_{X^*}^{-1}, && (153)\\
\mathbf{\Theta} &= \frac{a_W^2-\lambda_W}{2}\mathbf{\Omega}_{W^*}^{-1}, && (154)\\
c_X &= \frac{a_X^2-\lambda_X}{2}\log\left((2\pi)^n\left|\mathbf{\Omega}_{X^*}\right|\right)\\
c_Y &= 1-\frac{1}{2}\log\left((2\pi)^n\left|\mathbf{\Omega}_{X^*}+\mathbf{\Omega}_{W^*}\right|\right),\\
a_W^2-\lambda_W &= a_X^2-\lambda_X\geq 1, && (155)\\
\left|\mathbf{\Omega}_{X^*}\right| &= \left(\frac{1}{2\pi e}\exp\left\{\frac{2}{n}p_X\right\}\right)^n, && (156)\\
\left|\mathbf{\Omega}_{W^*}\right| &= \left(\frac{1}{2\pi e}\exp\left\{\frac{2}{n}p_W\right\}\right)^n. && (157)
\end{aligned}
$$

Without loss of generality, the matrices $\mathbf{\Omega}_{X^*}$ and $\mathbf{\Omega}_{W^*}$ are assumed to be invertible due to the same reasons mentioned in Appendix B. The relationships in (155) are obtained based on the second-order variation condition, which will be shown later in this proof.

Therefore, we can always find the Lagrange multipliers.

Now, consider the conditions for the second-order variation condition:

$$\begin{aligned}
K''_{f_X f_X}\Big|_{f_X=f_{X^*},f_W=f_{W^*},f_Y=f_{Y^*}} &= \frac{(a_X^2-\lambda_X)f_{W^*}(\mathbf{y}-\mathbf{x})}{f_{X^*}(\mathbf{x})}, \\
K''_{f_W f_W}\Big|_{f_X=f_{X^*},f_W=f_{W^*},f_Y=f_{Y^*}} &= \frac{(a_W^2-\lambda_W)f_{X^*}(\mathbf{x})}{f_{W^*}(\mathbf{y}-\mathbf{x})}, \\
\left(\int K d\mathbf{x}+\tilde{K}\right)''_{f_Y f_Y}\Big|_{f_X=f_{X^*},f_W=f_{W^*},f_Y=f_{Y^*}} &= \frac{f_{X^*}(\mathbf{x})f_{W^*}(\mathbf{y}-\mathbf{x})}{f_{Y^*}(\mathbf{y})^2}, \\
K''_{f_X f_W}\Big|_{f_X=f_{X^*},f_W=f_{W^*},f_Y=f_{Y^*}} &= a_W^2-\lambda_W, \\
K''_{f_W f_X}\Big|_{f_X=f_{X^*},f_W=f_{W^*},f_Y=f_{Y^*}} &= a_X^2-\lambda_X, \\
K''_{f_X f_Y}\Big|_{f_X=f_{X^*},f_W=f_{W^*},f_Y=f_{Y^*}} &= -\frac{f_{W^*}(\mathbf{y}-\mathbf{x})}{f_{Y^*}(\mathbf{y})}, \\
K''_{f_Y f_X}\Big|_{f_X=f_{X^*},f_W=f_{W^*},f_Y=f_{Y^*}} &= -\frac{f_{W^*}(\mathbf{y}-\mathbf{x})}{f_{Y^*}(\mathbf{y})}, \\
K''_{f_W f_Y}\Big|_{f_X=f_{X^*},f_W=f_{W^*},f_Y=f_{Y^*}} &= -\frac{f_{X^*}(\mathbf{x})}{f_{Y^*}(\mathbf{y})}, \\
K''_{f_Y f_W}\Big|_{f_X=f_{X^*},f_W=f_{W^*},f_Y=f_{Y^*}} &= -\frac{f_{X^*}(\mathbf{x})}{f_{Y^*}(\mathbf{y})}.
\end{aligned} \tag{158}$$

To satisfy $\delta^2 U[f_{X^*}, f_{W^*}, f_{Y^*}] \geq 0$, the following must hold:

$$\begin{aligned}
&\begin{bmatrix} h_X & h_W & h_Y \end{bmatrix} \begin{bmatrix} K''_{f_{X^*}f_{X^*}} & K''_{f_{X^*}f_{W^*}} & K''_{f_{X^*}f_{Y^*}} \\ K''_{f_{W^*}f_{X^*}} & K''_{f_{W^*}f_{W^*}} & K''_{f_{W^*}f_{Y^*}} \\ K''_{f_{Y^*}f_{X^*}} & K''_{f_{Y^*}f_{W^*}} & K''_{f_{Y^*}f_{Y^*}} \end{bmatrix} \begin{bmatrix} h_X \\ h_W \\ h_Y \end{bmatrix} \\
= \;& K''_{f_{X^*}f_{X^*}}h_X^2 + K''_{f_{W^*}f_{W^*}}h_W^2 + K''_{f_{Y^*}f_{Y^*}}h_Y^2 + (K''_{f_{X^*}f_{W^*}} + K''_{f_{W^*}f_{X^*}})h_X h_W \\
& +(K''_{f_{W^*}f_{Y^*}} + K''_{f_{Y^*}f_{W^*}})h_W h_Y + (K''_{f_{X^*}f_{Y^*}} + K''_{f_{Y^*}f_{X^*}})h_Y h_X \\
\geq \;& 0.
\end{aligned} \tag{159}$$

Using the defined quantities in (158), the equation in (159) is expressed as follows:

$$\begin{aligned}
& K''_{f_{X^*}f_{X^*}} h_X^2 + K''_{f_{W^*}f_{W^*}} h_W^2 + K''_{f_{Y^*}f_{Y^*}} h_Y^2 \\
& + (K''_{f_{X^*}f_{W^*}} + K''_{f_{W^*}f_{X^*}}) h_X h_W + (K''_{f_{W^*}f_{Y^*}} + K''_{f_{Y^*}f_{W^*}}) h_W h_Y + (K''_{f_{X^*}f_{Y^*}} + K''_{f_{Y^*}f_{X^*}}) h_Y h_X \\
= \; & \frac{(a_X^2 - \lambda_X) f_{W^*}(\mathbf{y}-\mathbf{x})}{f_{X^*}(\mathbf{x})} h_X(\mathbf{x})^2 + \frac{(a_W^2 - \lambda_W) f_{X^*}(\mathbf{x})}{f_{W^*(\mathbf{y}-\mathbf{x})}} h_W(\mathbf{y}-\mathbf{x})^2 + \frac{f_{X^*}(\mathbf{x}) f_{W^*}(\mathbf{y}-\mathbf{x})}{f_{Y^*}(\mathbf{y})^2} h_Y(\mathbf{y})^2 \\
& + 2\frac{(a_W^2 - \lambda_W)}{a_W} h_X(\mathbf{x}) h_W(\mathbf{y}-\mathbf{x}) - 2\frac{f_{X^*}(\mathbf{x})}{f_{Y^*}(\mathbf{y})} h_W(\mathbf{y}-\mathbf{x}) h_Y(\mathbf{y}) - 2\frac{f_{W^*}(\mathbf{y}-\mathbf{x})}{f_{Y^*}(\mathbf{y})} h_X(\mathbf{x}) h_Y(\mathbf{y}) \\
= \; & \frac{f_{W^*}(\mathbf{y}-\mathbf{x})}{f_{X^*}(\mathbf{x})} \Bigg((a_W^2 - \lambda_W) h_X(\mathbf{x})^2 + (a_W^2 - \lambda_W) \frac{f_{X^*}(\mathbf{x})^2}{f_{W^*}(\mathbf{y}-\mathbf{x})^2} h_W(\mathbf{y}-\mathbf{x})^2 + \frac{f_{X^*}(\mathbf{x})^2}{f_{Y^*}(\mathbf{y})^2} h_Y(\mathbf{y})^2 \\
& \qquad + 2(a_W^2 - \lambda_W) \frac{f_{X^*}(\mathbf{x})}{f_{W^*}(\mathbf{y}-\mathbf{x})} h_X(\mathbf{x}) h_W(\mathbf{y}-\mathbf{x}) \\
& \qquad - 2\frac{f_{X^*}(\mathbf{x})^2}{f_{W^*}(\mathbf{y}-\mathbf{x})\, f_{Y^*}(\mathbf{y})} h_W(\mathbf{y}-\mathbf{x}) h_Y(\mathbf{y}) - 2\frac{f_{X^*}(\mathbf{x})}{f_{Y^*}(\mathbf{y})} h_X(\mathbf{x}) h_Y(\mathbf{y}) \Bigg) \\
\geq \; & \frac{f_{W^*}(\mathbf{y}-\mathbf{x})}{f_{X^*}(\mathbf{x})} \left(h_X(\mathbf{x}) + \frac{f_{X^*}(\mathbf{x})}{f_{W^*}(\mathbf{y}-\mathbf{x})} h_W(\mathbf{y}-\mathbf{x}) - \frac{f_{X^*}(\mathbf{x})}{f_{Y^*}(\mathbf{y})} h_Y(\mathbf{y}) \right)^2 \\
\geq \; & 0,
\end{aligned} \tag{160}$$

where $a_W^2 - \lambda_W = a_X^2 - \lambda_X \geq 1$.

Therefore, the optimal solutions, f_{X^*}, f_{W^*}, and f_{Y^*}, minimize the variational problem in (146). Even though f_{X^*}, f_{W^*}, and f_{Y^*} are necessarily minimum solutions, multi-variate Gaussian density functions are the only ones in the feasible set. However, unlike Theorem 13, the correlation matrices are not explicitly defined as shown in (156) and (157), and there are more than one Gaussian density functions which satisfy the first-order and the second-order variation conditions. Therefore, we need an additional step to determine the correlation matrices $\mathbf{\Omega}_{X^*}$ and $\mathbf{\Omega}_{W^*}$ as follows.

Based on the first-order and the second-order variation conditions, we know the optimal solutions of the functional problem in (146) are multi-variate Gaussian density functions f_{X^*} and f_{W^*} whose correlation matrices are $\mathbf{\Omega}_{X^*}$ and $\mathbf{\Omega}_{W^*}$, respectively. Therefore, the inequality in (37) is expressed as

$$\begin{aligned}
& h(a_X\mathbf{X} + a_W\mathbf{W}) - a_X^2 h(\mathbf{X}) - a_W^2 h(\mathbf{W}) \\
\geq \; & h(a_X\mathbf{X}^* + a_W\mathbf{W}^*) - a_X^2 h(\mathbf{X}^*) - a_W^2 h(\mathbf{W}^*) \\
= \; & \frac{1}{2}\log(2\pi e)^n \left| a_X^2 \mathbf{\Omega}_{X^*} + a_W^2 \mathbf{\Omega}_{W^*} \right| - \frac{a_X^2}{2}\log(2\pi e)^n \left| \mathbf{\Omega}_{X^*} \right| - \frac{a_W^2}{2}\log(2\pi e)^n \left| \mathbf{\Omega}_{W^*} \right| \\
\geq \; & 0.
\end{aligned} \tag{161}$$

Since $\log|\cdot|$ is a concave function and $a_X^2 + a_W^2 = 1$, the inequality in (161) is proved using Jensen's

inequality. Therefore,

$$h(a_X\mathbf{X} + a_W\mathbf{W}) \geq a_X^2 h(\mathbf{X}) + a_W^2 h(\mathbf{W}), \tag{162}$$

and the proof is completed.

Remark 12: In (161), equality holds if and only if $\mathbf{\Omega}_{X^*} = \mathbf{\Omega}_{W^*}$. Since the optimal multi-variate Gaussian density functions have zero mean vectors, in this case, the correlation matrices are equal to the covariance matrices. Therefore, the equality condition requires identical covariance matrices. However, the equality condition is not required in the proof of EPI.

■

APPENDIX L
PROOF OF THEOREM 15

Proof: Now, construct the following variational problem, which represents the inequality in (38) and required constraints, as follows:

$$\begin{aligned}
\min_{f_X, f_Y} \quad & \int\int f_X(x) f_W(y-x) \left(-\mu \log f_Y(y) + \log f_X(x) + \mu\,(\mu - 1) \log f_W(y-x)\right) dxdy && (163)\\
\text{s.t.} \quad & \int\int f_X(x) f_W(y-x) dxdy = 1, \\
& \int\int (y-\mu_Y)^2 f_X(x) f_W(y-x) dxdy = \sigma_{Y^*}^2, \\
& \int\int (y-\mu_Y)^2 f_X(x) f_W(y-x) dxdy = \int\int (x-\mu_X)^2 f_X(x) f_W(y-x) dxdy \\
& \qquad\qquad + \int\int (y-x-\mu_W)^2 f_X(x) f_W(y-x) dxdy, \\
& \int\int (x-\mu_X)^2 f_X(x) f_W(y-x) dxdy \leq r^2, \\
& -\int\int f_X(x) f_W(y-x) \log f_X(x) dxdy = p, \\
& f_Y(y) = \int\int f_X(x) f_W(y-x) dxdy, && (164)
\end{aligned}$$

where p and r are constants, and $\sigma_{Y^*}^2$ stands for the variance of the optimal solution Y.

Using Lagrange multipliers, the functional problem in (163) is expressed as

$$\min_{f_X, f_Y} \quad \int \left(\int K(x, y, f_X, f_Y) dx \right) + \tilde{K}(y, f_Y) dy, \tag{165}$$

where

$$
\begin{aligned}
K(x,y,f_X,f_Y) &= f_X(x)f_W(y-x)\Big(-\mu\log f_Y(y)+\log f_X(x)+\mu(\mu-1)\log f_W(y-x)+\alpha_0 \\
&\quad +\beta_1(y-\mu_Y)^2+\beta_2(y-\mu_Y)^2-\beta_2(x-\mu_X)^2-\beta_2(y-x-\mu_W)^2+\beta_3(x-\mu_X)^2 \\
&\quad -\gamma_1\log f_X(x)-\lambda(y)\Big), \\
\tilde{K}(y,f_Y) &= \lambda(y)f_Y(y). \qquad (166)
\end{aligned}
$$

Due to the first-order variation condition,

$$
\begin{aligned}
& K'_{f_X}\Big|_{f_X=f_{X^*},f_Y=f_{Y^*}} \\
={}& f_W(y-x)\Big(-\mu\log f_{Y^*}(y)+\log f_{X^*}(x)+\mu(\mu-1)\log f_W(y-x)+\alpha_0 \\
& +\beta_1(y-\mu_Y)^2+\beta_2(y-\mu_Y)^2-\beta_2(x-\mu_X)^2-\beta_2(y-x-\mu_W)^2 \\
& +\beta_3(x-\mu_X)^2-\gamma_1\log f_{X^*}(x)-\lambda(y)+1-\gamma_1\Big) \\
={}& 0, \qquad (167) \\
& \int K'_{f_Y}dx+\tilde{K}'_{f_Y}\Big|_{f_X=f_{X^*},f_Y=f_{Y^*}} \\
={}& -\mu\frac{\int f_{X^*}(x)f_W(y-x)dx}{f_{Y^*}(y)}+\lambda(y) \\
={}& 0. \qquad (168)
\end{aligned}
$$

Since the equations in (167) and (168) must be satisfied for any x and y,

$$\begin{aligned}
\lambda(y) &= \mu, \\
f_{Y^*}(y) &= \exp\left\{\frac{1}{\mu}\left((\beta_1+\beta_2)(y-\mu_{Y^*})^2+c_Y\right)\right\} \\
&= \frac{1}{\sqrt{2\pi\left(-\frac{\mu}{2(\beta_1+\beta_2)}\right)}}\exp\left\{-\frac{1}{2\left(-\frac{\mu}{2(\beta_1+\beta_2)}\right)}(y-\mu_{Y^*})^2\right\} \\
&\quad \times\sqrt{2\pi\left(-\frac{\mu}{2(\beta_1+\beta_2)}\right)}\exp\left\{\frac{c_Y}{\mu}\right\} \\
f_W(y-x) &= \exp\left\{\frac{\beta_2}{\mu(\mu-1)}(y-x-\mu_W)^2-\frac{c_W}{\mu(\mu-1)}\right\} \\
&= \frac{1}{\sqrt{2\pi\left(-\frac{\mu(\mu-1)}{2(\beta_2)}\right)}}\exp\left\{-\frac{1}{2\left(-\frac{\mu(\mu-1)}{2(\beta_2)}\right)}(y-x-\mu_W)^2\right\} \\
&\quad \times\sqrt{2\pi\left(-\frac{\mu(\mu-1)}{2(\beta_2)}\right)}\exp\left\{-\frac{c_W}{\mu(\mu-1)}\right\}, \\
f_{X^*}(x) &= \exp\left\{\frac{1}{1-\gamma_1}\left((\beta_2-\beta_3)(x-\mu_{X^*})^2-\alpha_0+\mu-1+\gamma_1+c_W+c_Y\right)\right\} \\
&= \frac{1}{\sqrt{2\pi\left(-\frac{1-\gamma_1}{2(\beta_2-\beta_3)}\right)}}\exp\left\{-\frac{1}{2\left(-\frac{1-\gamma_1}{2(\beta_2-\beta_3)}\right)}(x-\mu_{X^*})^2\right\} \\
&\quad \times\sqrt{2\pi\left(-\frac{1-\gamma_1}{2(\beta_2-\beta_3)}\right)}\exp\left\{\frac{-\alpha_0+\mu-1+\gamma_1+c_W+c_Y}{1-\gamma_1}\right\}.
\end{aligned} \tag{169}$$

Considering the constraints in (164), the equations in (169) are further processed as follows:

$$
\begin{aligned}
f_{Y^*}(y) &= \frac{1}{\sqrt{2\pi\left(-\frac{\mu}{2(\beta_1+\beta_2)}\right)}} \exp\left\{-\frac{1}{2\left(-\frac{\mu}{2(\beta_1+\beta_2)}\right)}\left(y-\mu_{Y^*}\right)^2\right\} \\
&\quad \times\sqrt{2\pi\left(-\frac{\mu}{2\left(\beta_1+\beta_2\right)}\right)} \exp\left\{\frac{c_Y}{\mu}\right\} \\
&= \frac{1}{\sqrt{2\pi\sigma_{Y^*}^2}} \exp\left\{-\frac{1}{2\sigma_{Y^*}^2}\left(y-\mu_{Y^*}\right)^2\right\}, \\
f_W(y-x) &= \frac{1}{\sqrt{2\pi\left(-\frac{\mu(\mu-1)}{2(\beta_2)}\right)}} \exp\left\{-\frac{1}{2\left(-\frac{\mu(\mu-1)}{2(\beta_2)}\right)}\left(y-x-\mu_W\right)^2\right\} \\
&\quad \times\sqrt{2\pi\left(-\frac{\mu\left(\mu-1\right)}{2\left(\beta_2\right)}\right)} \exp\left\{-\frac{c_W}{\mu\left(\mu-1\right)}\right\} \\
&= \frac{1}{\sqrt{2\pi\sigma_W^2}} \exp\left\{-\frac{1}{2\sigma_W^2}\left(y-x-\mu_W\right)^2\right\}, \\
f_{X^*}(x) &= \frac{1}{\sqrt{2\pi\left(-\frac{1-\gamma_1}{2(\beta_2-\beta_3)}\right)}} \exp\left\{-\frac{1}{2\left(-\frac{1-\gamma_1}{2(\beta_2-\beta_3)}\right)}\left(x-\mu_{X^*}\right)^2\right\} \\
&\quad \times\sqrt{2\pi\left(-\frac{1-\gamma_1}{2\left(\beta_2-\beta_3\right)}\right)} \exp\left\{\frac{-\alpha_0+\mu-1+\gamma_1+c_W+c_Y}{1-\gamma_1}\right\} \\
&= \frac{1}{\sqrt{2\pi\sigma_{X^*}^2}} \exp\left\{-\frac{1}{2\sigma_{X^*}^2}\left(x-\mu_{X^*}\right)^2\right\}, \qquad (170)
\end{aligned}
$$

where

$$
\begin{aligned}
\alpha_0 &= \mu - (1-\gamma_1) + c_{W} + c_{Y} + \frac{1-\gamma_1}{2}\log\left(2\pi\sigma_{X^*}^2\right) \\
&= \frac{\mu(\mu-1)}{2}\log\left(2\pi m_W^2\right) - \frac{\mu}{2}\log\left(2\pi m_Y^2\right) + \frac{\mu}{2}\log\left(2\pi m_X^2\right), \\
\beta_1 &= -\beta_2 - \frac{\mu}{2\sigma_{Y^*}^2} \\
&= \frac{\mu(\mu-1)}{2\sigma_W^2} - \frac{\mu}{2\sigma_{Y^*}^2} \\
\beta_2 &= -\frac{\mu(\mu-1)}{2\sigma_W^2}, \\
\beta_3 &= \beta_2 + \frac{(1-\gamma_1)}{2\sigma_{X^*}^2} \\
&= -\frac{\mu(\mu-1)}{2\sigma_W^2} + \frac{(1-\gamma_1)}{2\sigma_{X^*}^2} \qquad (171)\\
&\geq 0, \\
c_W &= \frac{\mu(\mu-1)}{2}\log\left(2\pi\sigma_W^2\right), \\
c_Y &= -\frac{\mu}{2}\log\left(2\pi\sigma_{Y^*}^2\right), \\
\sigma_{X^*}^2 &= \frac{1}{2\pi e}\exp\{2p\} \leq r^2, \qquad (172)\\
\sigma_{Y^*}^2 &= \sigma_{X^*}^2 + \sigma_W^2, \\
\gamma_1 &\leq 1-\mu. \qquad (173)
\end{aligned}
$$

The constant p must be chosen to satisfy the inequality in (172) due to Theorem 4. The inequality in (173) is due to the second-order variation condition, which will be presented later in this proof. Therefore, by appropriately choosing p, the Lagrange multipliers always exist, and therefore, the necessary optimal solutions, which are Gaussian, exist.

To make the second variation positive, we need the positive-definiteness of the following matrix:

$$
\left.\begin{bmatrix} K''_{f_X f_X} & K''_{f_X f_Y} \\ K''_{f_Y f_X} & K''_{f_Y f_Y} \end{bmatrix}\right|_{f_X = f_{X^*}, f_Y = f_{Y^*}} \qquad (174)
$$

and it requires the following:

$$
\begin{aligned}
& \left.\begin{bmatrix} h_X & h_Y \end{bmatrix}\begin{bmatrix} K''_{f_X f_X} & K''_{f_X f_Y} \\ K''_{f_Y f_X} & K''_{f_Y f_Y} \end{bmatrix}\begin{bmatrix} h_X \\ h_Y \end{bmatrix}\right|_{f_X = f_{X^*}, f_Y = f_{Y^*}} \\
= {} & \left. K''_{f_X f_X} h_X^2 + K''_{f_Y f_Y} h_Y^2 + (K''_{f_X f_Y} + K''_{f_Y f_X}) h_Y h_X \right|_{f_X = f_{X^*}, f_Y = f_{Y^*}} \qquad (175)\\
> {} & 0,
\end{aligned}
$$

where h_X and h_Y are arbitrary admissible functions.

Since $K''_{f_X f_X}$, $K''_{f_X f_Y}$, $K''_{f_Y f_X}$, and $K''_{f_Y f_Y}$ are defined as

$$\begin{aligned} K''_{f_X f_X} &= \frac{(1-\gamma_1) f_W(y-x)}{f_X(x)}, \\ K''_{f_X f_Y} &= -\frac{\mu f_W(y-x)}{f_Y(y)}, \\ K''_{f_Y f_X} &= -\frac{\mu f_W(y-x)}{f_Y(y)}, \\ K''_{f_Y f_Y} &= \frac{\mu f_X(x) f_W(y-x)}{f_Y(y)^2}, \end{aligned} \tag{176}$$

the equation in (175) requires the following:

$$\begin{aligned} &\frac{(1-\gamma_1) f_W(y-x)}{f_{X^*}(x)} h_X(x)^2 - 2\frac{\mu f_W(y-x)}{f_{Y^*}(y)} h_X(x) h_Y(y) + \frac{\mu f_{X^*}(x) f_W(y-x)}{f_{Y^*}(y)^2} h_Y(y)^2 \\ \geq\ & \frac{\mu f_W(y-x)}{f_{X^*}(x)} \left(h_X(x) - \frac{f_{X^*}(x)}{f_{Y^*}(y)} h_Y(y) \right)^2, \end{aligned} \tag{177}$$

where $\gamma_1 \leq 1-\mu$. Similar to the complementary slackness in KKT conditions, when $\beta_3 = 0$ in (171), $\sigma^2_{X^*} = (1-\gamma_1)\,\mu^{-1}(\mu-1)^{-1}\sigma^2_{W^*}$, and it requires $(1-\gamma_1)\,\mu^{-1}(\mu-1)^{-1}\sigma^2_{W^*} < r^2$ (If $\gamma_1 = 1-\mu$, then $\sigma^2_{X^*} = (\mu-1)^{-1}\sigma^2_{W^*}$). Otherwise, $\sigma^2_{X^*} = r^2 \leq (1-\gamma_1)\,\mu^{-1}(\mu-1)^{-1}\sigma^2_{W^*}$.

In conclusion, the Gaussian density function, whose variance is $\sigma^2_{X^*}$, minimizes the variational problem in (163), and the proof is completed.

Remark 13: Unlike other theorems shown in this paper, Theorem 15 only requires to find necessarily optimal solutions, a result similar to Theorem 8 in [2].

■

APPENDIX M
PROOF OF THEOREM 16

Proof: We first construct the following variational problem (without loss of generality, we assume the mean vectors of $\mathbf{X}$, $\mathbf{W}$, and $\mathbf{Y}$ are zeros. (cf. Appendix L)):

$$\min_{f_X, f_Y} \quad \int\int f_X(\mathbf{x}) f_W(\mathbf{y}-\mathbf{x}) \left(-\mu \log f_Y(\mathbf{y}) + \log f_X(\mathbf{x}) + \mu(\mu-1) \log f_W(\mathbf{y}-\mathbf{x})\right) d\mathbf{x} d\mathbf{y} \quad (178)$$

$$\text{s.t.} \quad \int\int f_X(\mathbf{x}) f_W(\mathbf{y}-\mathbf{x}) d\mathbf{x} d\mathbf{y} = 1,$$

$$\int\int \mathbf{y}\mathbf{y}^T f_X(\mathbf{x}) f_W(\mathbf{y}-\mathbf{x}) d\mathbf{x} d\mathbf{y} = \int\int \mathbf{x}\mathbf{x}^T f_X(\mathbf{x}) f_W(\mathbf{y}-\mathbf{x}) d\mathbf{x} d\mathbf{y},$$

$$+ \int\int (\mathbf{y}-\mathbf{x})(\mathbf{y}-\mathbf{x})^T f_X(\mathbf{x}) f_W(\mathbf{y}-\mathbf{x}) d\mathbf{x} d\mathbf{y},$$

$$\int\int \mathbf{x}\mathbf{x}^T f_X(\mathbf{x}) f_W(\mathbf{y}-\mathbf{x}) d\mathbf{x} d\mathbf{y} \preceq \mathbf{\Sigma},$$

$$\int\int \mathbf{y}\mathbf{y}^T f_X(\mathbf{x}) f_W(\mathbf{y}-\mathbf{x}) d\mathbf{x} d\mathbf{y} = \mathbf{\Sigma}_{Y^*},$$

$$-\int\int f_X(\mathbf{x}) f_W(\mathbf{y}-\mathbf{x}) \log f_X(\mathbf{x}) d\mathbf{x} d\mathbf{y} = p_X,$$

$$f_Y(\mathbf{y}) = \int\int f_X(\mathbf{x}) f_W(\mathbf{y}-\mathbf{x}) d\mathbf{x} d\mathbf{y}, \quad (179)$$

where p_X is a constant, and $\mathbf{\Sigma}_{Y^*}$ is the covariance matrix of the optimal solution of $\mathbf{Y}$. Without loss of generality, the matrix $\mathbf{\Sigma}$ is assumed to be a positive definite matrix due to the same reason mentioned in [2].

This problem is more appropriately changed as follows:

$$\min_{f_X, f_Y} \quad \int\int f_X(\mathbf{x}) f_W(\mathbf{y}-\mathbf{x}) \left(-\mu \log f_Y(\mathbf{y}) + \log f_X(\mathbf{x}) + \mu(\mu-1) \log f_W(\mathbf{y}-\mathbf{x})\right) d\mathbf{x} d\mathbf{y} \quad (180)$$

$$\text{s.t.} \quad \int\int f_X(\mathbf{x}) f_W(\mathbf{y}-\mathbf{x}) d\mathbf{x} d\mathbf{y} = 1,$$

$$\int\int \left(y_i y_j - x_i x_j - (y-x)_i (y-x)_j\right) f_X(\mathbf{x}) f_W(\mathbf{y}-\mathbf{x}) d\mathbf{x} d\mathbf{y} = 0,$$

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \left(\int\int x_i x_j \xi_i \xi_j f_X(\mathbf{x}) f_W(\mathbf{y}-\mathbf{x}) d\mathbf{x} d\mathbf{y} \right) \leq \sum_{i=1}^{n} \sum_{j=1}^{n} \sigma_{ij}^2 \xi_i \xi_j,$$

$$\int\int y_i y_j f_X(\mathbf{x}) f_W(\mathbf{y}-\mathbf{x}) d\mathbf{x} d\mathbf{y} = \sigma_{Y_{ij}^*}^2,$$

$$-\int\int f_X(\mathbf{x}) f_W(\mathbf{y}-\mathbf{x}) \log f_X(\mathbf{x}) d\mathbf{x} d\mathbf{y} = p_X,$$

$$f_Y(\mathbf{y}) = \int\int f_X(\mathbf{x}) f_W(\mathbf{y}-\mathbf{x}) d\mathbf{x} d\mathbf{y}, \quad (181)$$

where the arbitrary deterministic non-zero vector $\boldsymbol{\xi}$ is defined as $[\xi_1, \ldots, \xi_n]^T$, $\sigma^2_{Y^*_{ij}}$ denotes the i^{th} row and j^{th} column element of $\boldsymbol{\Sigma}_{Y^*}$, $i = 1, \ldots, n$, and $j = 1, \ldots, n$.

Using Lagrange multipliers, the functional problem in (180) and the constraints in (181) are expressed as

$$\min_{f_X, f_Y} \quad \int \left(\int K(\mathbf{x}, \mathbf{y}, f_X, f_Y) d\mathbf{x} \right) + \tilde{K}(\mathbf{y}, f_Y) d\mathbf{y}, \tag{182}$$

where

$$\begin{aligned} K(\mathbf{x}, \mathbf{y}, f_X, f_Y) &= f_X(\mathbf{x}) f_W(\mathbf{y}-\mathbf{x}) \Big(-\mu \log f_Y(\mathbf{y}) + \log f_X(\mathbf{x}) + \mu\,(\mu - 1) \log f_W(\mathbf{y}-\mathbf{x}) + \alpha_0 \\ &\quad + \sum_{i=1}^{n} \sum_{j=1}^{n} \Big(\gamma_{ij} y_i y_j - \gamma_{ij} x_i x_j - \gamma_{ij} \,(y-x)_i\, (y-x)_j + \theta x_i x_j \xi_i \xi_j + \phi_{ij} y_i y_j \Big) \\ &\quad - \alpha_1 \log f_X(\mathbf{x}) - \lambda(\mathbf{y}) \Big), \\ \tilde{K}(\mathbf{y}, f_Y) &= \lambda(\mathbf{y}) f_Y(\mathbf{y}). \end{aligned} \tag{183}$$

Then, the first-order variation condition is checked as follows.

$$\begin{aligned} K'_{f_X}\Big|_{f_X = f_{X^*}, f_Y = f_{Y^*}} &= f_W(\mathbf{y}-\mathbf{x}) \Big(-\mu \log f_{Y^*}(\mathbf{y}) + (1-\alpha_1) \log f_{X^*}(\mathbf{x}) \\ &\quad + \mu\,(\mu-1) \log f_W(\mathbf{y}-\mathbf{x}) + \alpha_0 + \sum_{i=1}^{n} \sum_{j=1}^{n} \Big(\gamma_{ij} y_i y_j - \gamma_{ij} x_i x_j \\ &\quad - \gamma_{ij}\,(y-x)_i\,(y-x)_j + \theta x_i x_j \xi_i \xi_j + \phi_{ij} y_i y_j + \Big) - \lambda(\mathbf{y}) + 1 - \alpha_1 \Big) \\ &= 0. \end{aligned} \tag{184}$$

$$\begin{aligned} K'_{f_X}\Big|_{f_Y = f_{X^*}, f_Y = f_{Y^*}} &= -\frac{\mu \int f_X(\mathbf{x}) f_W(\mathbf{y}-\mathbf{x}) d\mathbf{x}}{f_Y(\mathbf{y})} + \lambda(\mathbf{y}) \\ &= 0. \end{aligned} \tag{185}$$

Since the equalities in (184) and (185) must be satisfied for any $\mathbf{x}$ and $\mathbf{y}$,

$$\begin{aligned}
\lambda(\mathbf{y}) &= \mu, \\
f_{Y^*}(\mathbf{y}) &= \exp\left\{\frac{1}{\mu}\left(\mathbf{y}^T\left(\mathbf{\Gamma}+\mathbf{\Phi}\right)\mathbf{y}+c_Y\right)\right\} \\
&= (2\pi)^{-\frac{n}{2}}\left|-\frac{\mu}{2}\left(\mathbf{\Gamma}+\mathbf{\Phi}\right)^{-1}\right|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}\mathbf{y}^T\left(-\frac{\mu}{2}\left(\mathbf{\Gamma}+\mathbf{\Phi}\right)^{-1}\right)^{-1}\mathbf{y}\right\} \\
&\quad\times(2\pi)^{\frac{n}{2}}\left|-\frac{\mu}{2}\left(\mathbf{\Gamma}+\mathbf{\Phi}\right)^{-1}\right|^{\frac{1}{2}}\exp\left\{\frac{c_Y}{\mu}\right\} \\
f_W(\mathbf{y}-\mathbf{x}) &= \exp\left\{\frac{1}{\mu\left(\mu-1\right)}\left(\left(\mathbf{y}-\mathbf{x}\right)^T\mathbf{\Gamma}\left(\mathbf{y}-\mathbf{x}\right)-c_W\right)\right\} \\
&= (2\pi)^{-\frac{n}{2}}\left|-\frac{\mu\left(\mu-1\right)}{2}\mathbf{\Gamma}^{-1}\right|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}\left(\mathbf{y}-\mathbf{x}\right)^T\left(-\frac{\mu\left(\mu-1\right)}{2}\mathbf{\Gamma}^{-1}\right)^{-1}\left(\mathbf{y}-\mathbf{x}\right)\right\} \\
&\quad\times(2\pi)^{\frac{n}{2}}\left|-\frac{\mu\left(\mu-1\right)}{2}\mathbf{\Gamma}^{-1}\right|^{\frac{1}{2}}\exp\left\{-\frac{c_W}{\mu\left(\mu-1\right)}\right\}, \\
f_{X^*}(\mathbf{x}) &= \exp\left\{\frac{1}{1-\alpha_1}\left(\mathbf{x}^T\left(\mathbf{\Gamma}-\theta\mathbf{\Xi}\right)\mathbf{x}-\alpha_0+\mu-1+\alpha_1+c_W+c_Y\right)\right\} \\
&= (2\pi)^{-\frac{n}{2}}\left|-\frac{1-\alpha_1}{2}\left(\mathbf{\Gamma}-\theta\mathbf{\Xi}\right)^{-1}\right|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}\mathbf{x}^T\left(-\frac{1-\alpha_1}{2}\left(\mathbf{\Gamma}-\theta\mathbf{\Xi}\right)^{-1}\right)^{-1}\mathbf{x}\right\} \\
&\quad\times(2\pi)^{\frac{n}{2}}\left|-\frac{1-\alpha_1}{2}\left(\mathbf{\Gamma}-\theta\mathbf{\Xi}\right)^{-1}\right|^{\frac{1}{2}}\exp\left\{\frac{-\alpha_0+\mu-1+\alpha_1+c_W+c_Y}{1-\alpha_1}\right\}, \qquad (186)
\end{aligned}$$

where

$$\begin{aligned}
\mathbf{\Phi} &= \begin{bmatrix}\phi_{11} & \cdots & \phi_{1n}\\ \vdots & \ddots & \vdots \\ \phi_{n1} & \cdots & \phi_{nn}\end{bmatrix}, \quad \mathbf{\Gamma} = \begin{bmatrix}\gamma_{11} & \cdots & \gamma_{1n}\\ \vdots & \ddots & \vdots \\ \gamma_{n1} & \cdots & \gamma_{nn}\end{bmatrix}, \quad \mathbf{\Xi} = \begin{bmatrix}\xi_1\xi_1 & \cdots & \xi_1\xi_n\\ \vdots & \ddots & \vdots \\ \xi_n\xi_1 & \cdots & \xi_n\xi_n\end{bmatrix}, \\
\mathbf{x} &= \left[x_1,\cdots,x_n\right]^T, \\
\mathbf{y} &= \left[y_1,\cdots,y_n\right]^T, \\
\theta &\geq 0. \qquad (187)
\end{aligned}$$

Considering the constraints in (181), the equations in (186) are further processed as follows.

$$\begin{aligned}
f_{Y^*}(y) &= (2\pi)^{-\frac{n}{2}} \left| -\frac{\mu}{2} (\mathbf{\Gamma} + \mathbf{\Phi})^{-1} \right|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} \mathbf{y}^T \left(-\frac{\mu}{2} (\mathbf{\Gamma} + \mathbf{\Phi})^{-1} \right)^{-1} \mathbf{y} \right\} \\
&\quad \times (2\pi)^{\frac{n}{2}} \left| -\frac{\mu}{2} (\mathbf{\Gamma} + \mathbf{\Phi})^{-1} \right|^{\frac{1}{2}} \exp\left\{ \frac{c_Y}{\mu} \right\} \\
&= (2\pi)^{-\frac{n}{2}} \left| \mathbf{\Sigma}_{Y^*} \right|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} \mathbf{y}^T \mathbf{\Sigma}_{Y^*}^{-1} \mathbf{y} \right\}, \\
f_W(y - x) &= (2\pi)^{-\frac{n}{2}} \left| -\frac{\mu(\mu - 1)}{2} \mathbf{\Gamma}^{-1} \right|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} (\mathbf{y} - \mathbf{x})^T \left(-\frac{\mu(\mu - 1)}{2} \mathbf{\Gamma}^{-1} \right)^{-1} (\mathbf{y} - \mathbf{x}) \right\} \\
&\quad \times (2\pi)^{\frac{n}{2}} \left| -\frac{\mu(\mu - 1)}{2} \mathbf{\Gamma}^{-1} \right|^{\frac{1}{2}} \exp\left\{ -\frac{c_W}{\mu(\mu - 1)} \right\} \\
&= (2\pi)^{-\frac{n}{2}} \left| \mathbf{\Sigma}_W \right|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} (\mathbf{y} - \mathbf{x})^T \mathbf{\Sigma}_W^{-1} (\mathbf{y} - \mathbf{x}) \right\}, \\
f_{X^*}(x) &= (2\pi)^{-\frac{n}{2}} \left| -\frac{1 - \alpha_1}{2} (\mathbf{\Gamma} - \theta \mathbf{\Xi})^{-1} \right|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} \mathbf{x}^T \left(-\frac{1 - \alpha_1}{2} (\mathbf{\Gamma} - \theta \mathbf{\Xi})^{-1} \right)^{-1} \mathbf{x} \right\} \\
&\quad \times (2\pi)^{\frac{n}{2}} \left| -\frac{1 - \alpha_1}{2} (\mathbf{\Gamma} - \theta \mathbf{\Xi})^{-1} \right|^{\frac{1}{2}} \exp\left\{ \frac{-\alpha_0 + \mu - 1 + \alpha_1 + c_W + c_Y}{1 - \alpha_1} \right\} \\
&= (2\pi)^{-\frac{n}{2}} \left| \mathbf{\Sigma}_{X^*} \right|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} \mathbf{x}^T \mathbf{\Sigma}_{X^*}^{-1} \mathbf{x} \right\},
\end{aligned} \tag{188}$$

where

$$
\begin{aligned}
\alpha_0 &= \mu - (1-\alpha_1) + c_W + c_Y + \frac{1-\alpha_1}{2}\log(2\pi)^n |\mathbf{\Sigma}_{X^*}| \\
&= \mu - (1-\alpha_1) + \frac{\mu(\mu-1)}{2}\log(2\pi)^n |\mathbf{\Sigma}_W| - \frac{\mu}{2}\log(2\pi)^n |\mathbf{\Sigma}_{Y^*}| + \frac{1-\alpha_1}{2}\log(2\pi)^n |\mathbf{\Sigma}_{X^*}|, \\
\mathbf{\Gamma} &= -\frac{\mu(\mu-1)}{2}\mathbf{\Sigma}_W^{-1}, \\
\mathbf{\Phi} &= -\mathbf{\Gamma} - \frac{\mu}{2}\mathbf{\Sigma}_{Y^*}^{-1} \\
&= \frac{\mu(\mu-1)}{2}\mathbf{\Sigma}_W^{-1} - \frac{\mu}{2}(\mathbf{\Sigma}_{X^*} + \mathbf{\Sigma}_W)^{-1}, \\
\mathbf{\Sigma}_{X^*} &= -\frac{1-\alpha_1}{2}(\mathbf{\Gamma} - \theta\mathbf{\Xi})^{-1} \\
&= \frac{1-\alpha_1}{2}\left(\frac{\mu(\mu-1)}{2}\mathbf{\Sigma}_W^{-1} + \theta\mathbf{\Xi}\right)^{-1} && (189) \\
&\succeq \mathbf{0}, && (190) \\
\theta &\geq 0, \\
\alpha_1 &\leq 1-\mu, && (191) \\
c_W &= \frac{\mu(\mu-1)}{2}\log(2\pi)^n |\mathbf{\Sigma}_W|, \\
c_Y &= -\frac{\mu}{2}\log(2\pi)^n |\mathbf{\Sigma}_{Y^*}|, \\
|\mathbf{\Sigma}_{X^*}| &= \left(\frac{1}{2\pi e}\exp\left\{\frac{2}{n}p_X\right\}\right)^n \leq |\mathbf{\Sigma}|. && (192)
\end{aligned}
$$

The inequality in (190) is always satisfied since the matrix $\mathbf{\Xi}$ is non-zero positive semi-definite and θ is non-negative. The inequality in (192) will be proved later in this proof. The constant p_X must be chosen to satisfy the inequality in (192). Then, the Lagrange multipliers always exist, and necessary optimal solutions exist.

Interestingly, similar to the complementary slackness in KKT conditions, when $\theta = 0$ in (189), $\mathbf{\Sigma}_{X^*} = (1-\alpha_1)\mu^{-1}(\mu-1)^{-1}\mathbf{\Sigma}_W$, and it requires $(1-\alpha_1)\mu^{-1}(\mu-1)^{-1}\mathbf{\Sigma}_W \preceq \mathbf{\Sigma}$. When θ is non-zero, the equation in (189) is positive semi-definite, and it means $\mathbf{\Sigma}_{X^*} = (1-\alpha_1)\mu^{-1}(\mu-1)^{-1}\mathbf{\Sigma}_{\tilde{W}}$, where $\mathbf{\Sigma}_{\tilde{W}} = \mathbf{\Sigma}_W - \mathbf{\Sigma}_{\hat{W}}$, where $\mathbf{\Sigma}_{\hat{W}}$ and $\mathbf{\Sigma}_{\tilde{W}}$ are positive semi-definite matrices. When $1-\alpha_1 = \mu$, then $\mathbf{\Sigma}_{X^*} = (\mu-1)^{-1}\mathbf{\Sigma}_{\tilde{W}}$, which is exactly the same as the one in [2] and [27].

To make the second variation positive, we need the positive-definiteness of the following matrix:

$$
\begin{bmatrix} K''_{f_{X^*}f_{X^*}} & K''_{f_{X^*}f_{Y^*}} \\ K''_{f_{Y^*}f_{X^*}} & K''_{f_{Y^*}f_{Y^*}} \end{bmatrix}, \tag{193}
$$

and it requires the following condition to hold:

$$\begin{aligned}
&\begin{bmatrix} h_X & h_Y \end{bmatrix} \begin{bmatrix} K''_{f_{X^*}f_{X^*}} & K''_{f_{X^*}f_{Y^*}} \\ K''_{f_{Y^*}f_{X^*}} & K''_{f_{Y^*}f_{Y^*}} \end{bmatrix} \begin{bmatrix} h_X \\ h_Y \end{bmatrix} \\
= \;& K''_{f_{X^*}f_{X^*}} h_X^2 + K''_{f_{Y^*}f_{Y^*}} h_Y^2 + (K''_{f_{X^*}f_{Y^*}} + K''_{f_{Y^*}f_{X^*}}) h_Y h_X \\
\geq \;& 0,
\end{aligned} \tag{194}$$

where h_X and h_Y are arbitrary admissible functions.

Since $K''_{f_{X^*}f_{X^*}}$, $K''_{f_{X^*}f_{Y^*}}$, $K''_{f_{Y^*}f_{X^*}}$, and $K''_{f_{Y^*}f_{Y^*}}$ are defined as

$$\begin{aligned}
K''_{f_{X^*}f_{X^*}} &= \frac{(1-\alpha_1) f_W(\mathbf{y}-\mathbf{x})}{f_{X^*}(\mathbf{x})}, \\
K''_{f_{X^*}f_{Y^*}} &= -\frac{\mu f_W(\mathbf{y}-\mathbf{x})}{f_{Y^*}(\mathbf{y})}, \\
K''_{f_{Y^*}f_{X^*}} &= -\frac{\mu f_W(\mathbf{y}-\mathbf{x})}{f_{Y^*}(\mathbf{y})}, \\
K''_{f_{Y^*}f_{Y^*}} &= \frac{\mu f_{X^*}(\mathbf{x}) f_W(\mathbf{y}-\mathbf{x})}{f_{Y^*}(\mathbf{y})^2},
\end{aligned} \tag{195}$$

the equation in (194) requires

$$\begin{aligned}
&\frac{(1-\alpha_1) f_W(\mathbf{y}-\mathbf{x})}{f_{X^*}(\mathbf{x})} h_X(\mathbf{x})^2 - 2\frac{\mu f_W(\mathbf{y}-\mathbf{x})}{f_{Y^*}(\mathbf{y})} h_X(\mathbf{x}) h_Y(\mathbf{y}) + \frac{\mu f_{X^*}(\mathbf{x}) f_W(\mathbf{y}-\mathbf{x})}{f_{Y^*}(\mathbf{y})^2} h_Y(\mathbf{y})^2 \\
\geq \;& \frac{\mu f_W(\mathbf{y}-\mathbf{x})}{f_{X^*}(\mathbf{x})} \left(h_X(\mathbf{x}) - \frac{f_{X^*}(\mathbf{x})}{f_{Y^*}(\mathbf{y})} h_Y(\mathbf{y}) \right)^2,
\end{aligned} \tag{196}$$

where $\alpha_1 \geq 1-\mu$.

Therefore, the optimal solutions f_{X^*} and f_{Y^*} minimize the functional problem in (180), and the proof is completed. ■

APPENDIX N
PROOF OF THEOREM 17

Proof: First, choose a Gaussian random vector $\tilde{\mathbf{W}}_G$ whose covariance matrix $\mathbf{\Sigma}_{\tilde{W}}$ satisfies $\mathbf{\Sigma}_{\tilde{W}} \preceq \mathbf{\Sigma}_W$ and $\mathbf{\Sigma}_{\tilde{W}} \preceq \mathbf{\Sigma}_V$. Since the Gaussian random vectors $\mathbf{V}_G$ and $\mathbf{W}_G$ can be represented as the summation of two independent random vectors $\tilde{\mathbf{W}}_G$ and $\hat{\mathbf{V}}_G$, and the summation of two independent random vectors $\tilde{\mathbf{W}}_G$ and $\hat{\mathbf{W}}_G$, respectively, the left-hand side of the equation in (40) is written as follows:

$$\begin{aligned}
&\mu h(\mathbf{X}+\mathbf{V}_G) - h(\mathbf{X}+\mathbf{W}_G) \\
\geq \;& \mu h(\mathbf{X}+\mathbf{V}_G) - h(\mathbf{X}+\tilde{\mathbf{W}}_G) - h(\mathbf{W}_G) + h(\tilde{\mathbf{W}}_G) \\
= \;& \mu h(\mathbf{X}+\tilde{\mathbf{W}}_G+\hat{\mathbf{V}}_G) - h(\mathbf{X}+\tilde{\mathbf{W}}_G) - h(\tilde{\mathbf{W}}_G+\hat{\mathbf{W}}_G) + h(\tilde{\mathbf{W}}_G).
\end{aligned} \tag{197}$$

Since the equation will be minimized over $f_X(\mathbf{x})$, the last two terms in (197) are ignored, and by substituting $\mathbf{Y}$ and $\hat{\mathbf{X}}$ for $\mathbf{X}+\tilde{\mathbf{W}}_G+\hat{\mathbf{V}}_G$ and $\mathbf{X}+\tilde{\mathbf{W}}_G$, respectively, the inequality in (40) is equivalently expressed as the following variational problem:

$$\begin{aligned}
\min_{f_{\hat{X}},f_Y} \quad & \mu h(\mathbf{Y}) - h(\hat{\mathbf{X}}) - \mu\,(\mu-1)\,h(\hat{\mathbf{V}}_G) \\
\text{s. t.} \quad & \int\int f_{\hat{X}}(\mathbf{x}) f_{\hat{V}}(\mathbf{y}-\mathbf{x}) d\mathbf{x} d\mathbf{y} - 1 = 0, \\
& \int\int f_{\hat{X}}(\mathbf{x}) f_{\hat{V}}(\mathbf{y}-\mathbf{x}) \mathbf{x}\mathbf{x}^T d\mathbf{x} d\mathbf{y} - \mathbf{\Sigma}_{\hat{X}} \preceq \mathbf{0}, \\
& \int\int f_{\hat{X}}(\mathbf{x}) f_{\hat{V}}(\mathbf{y}-\mathbf{x}) \mathbf{y}\mathbf{y}^T d\mathbf{x} d\mathbf{y} - \mathbf{\Sigma}_{Y^*} = \mathbf{0}, \\
& \int\int f_{\hat{X}}(\mathbf{x}) f_{\hat{V}}(\mathbf{y}-\mathbf{x}) \left(\mathbf{y}\mathbf{y}^T - \mathbf{x}\mathbf{x}^T - (\mathbf{y}-\mathbf{x})(\mathbf{y}-\mathbf{x})^T\right) d\mathbf{x} d\mathbf{y} = \mathbf{0}, \\
& -\int\int f_{\hat{X}}(\mathbf{x}) f_{\hat{V}}(\mathbf{y}-\mathbf{x}) \log f_{\hat{X}}(\mathbf{x}) d\mathbf{x} d\mathbf{y} = p_{\hat{X}}, \\
& f_Y(\mathbf{y}) = \int f_{\hat{X}}(\mathbf{x}) f_{\hat{V}}(\mathbf{y}-\mathbf{x}) d\mathbf{x},
\end{aligned} \tag{198}$$

where $\hat{\mathbf{X}} = \mathbf{X} + \tilde{\mathbf{W}}_G$, $\mathbf{Y} = \hat{\mathbf{X}} + \hat{\mathbf{V}}_G$, $\mathbf{W}_G = \tilde{\mathbf{W}}_G + \hat{\mathbf{W}}_G$, $\mathbf{V}_G = \tilde{\mathbf{W}}_G + \hat{\mathbf{V}}_G$, $\mathbf{\Sigma}_{\hat{X}} = \mathbf{\Sigma} + \mathbf{\Sigma}_{\tilde{W}}$, $\mathbf{\Sigma}_{Y^*} = \mathbf{\Sigma}_{X^*} + \mathbf{\Sigma}_V$, and $\mathbf{\Sigma}_{X^*}$ is the covariance matrix of the optimal solution $\mathbf{X}^*$.

The variational problem in (198) is exactly the same as the one in (180). Therefore, using the same method as in the proof of Theorem 16, we obtain the following inequality (see the details of the proof in Appendix M):

$$\begin{aligned}
& \mu h(\mathbf{X}+\tilde{\mathbf{W}}_G+\hat{\mathbf{V}}_G) - h(\mathbf{X}+\tilde{\mathbf{W}}_G) - h(\tilde{\mathbf{W}}_G+\hat{\mathbf{W}}_G) + h(\tilde{\mathbf{W}}_G) \\
\geq \; & \mu h(\mathbf{X}_G^*+\tilde{\mathbf{W}}_G+\hat{\mathbf{V}}_G) - h(\mathbf{X}_G^*+\tilde{\mathbf{W}}_G) - h(\tilde{\mathbf{W}}_G+\hat{\mathbf{W}}_G) + h(\tilde{\mathbf{W}}_G).
\end{aligned} \tag{199}$$

By appropriately choosing $\mathbf{X}_G^*$ and $\tilde{\mathbf{W}}_G$, the right-hand side of the equation in (199) is expressed as

$$\begin{aligned}
& \mu h(\mathbf{X}_G^*+\tilde{\mathbf{W}}_G+\hat{\mathbf{V}}_G) - h(\mathbf{X}_G^*+\tilde{\mathbf{W}}_G) - h(\tilde{\mathbf{W}}_G+\hat{\mathbf{W}}_G) + h(\tilde{\mathbf{W}}_G) \\
= \; & \mu h(\mathbf{X}_G^*+\tilde{\mathbf{W}}_G+\hat{\mathbf{V}}_G) - h(\mathbf{X}_G^*+\mathbf{W}_G).
\end{aligned} \tag{200}$$

The equality in (200) is due to the equality condition of data processing inequality in [27]. For the completeness of the proof, we introduce a technique, which is slightly different from the one in [27].

To satisfy the equality in the equation (200), the equality condition in the following lemma must be satisfied.

Lemma 1 (Data Processing Inequality [1]): When three random vectors $\mathbf{Y}_1$, $\mathbf{Y}_2$, and $\mathbf{Y}_3$ represent a Markov chain $\mathbf{Y}_1 \to \mathbf{Y}_2 \to \mathbf{Y}_3$, the following inequality is satisfied:

$$I(\mathbf{Y}_1;\mathbf{Y}_3) \leq I(\mathbf{Y}_1;\mathbf{Y}_2). \tag{201}$$

The equality holds if and only if $I(\mathbf{Y}_1;\mathbf{Y}_2|\mathbf{Y}_3)=0$.

In Lemma 1, $\mathbf{Y}_1$, $\mathbf{Y}_2$, and $\mathbf{Y}_3$ are defined as $\mathbf{X}_G^*$, $\mathbf{X}_G^*+\tilde{\mathbf{W}}_G$, and $\mathbf{X}_G^*+\tilde{\mathbf{W}}_G+\hat{\mathbf{W}}_G$, respectively. Therefore, the equality condition, $I(\mathbf{Y}_1;\mathbf{Y}_2|\mathbf{Y}_3)=0$ is expressed as

$$\begin{aligned}
I(\mathbf{Y}_1;\mathbf{Y}_2|\mathbf{Y}_3) &= h(\mathbf{Y}_1|\mathbf{Y}_3)-h(\mathbf{Y}_1|\mathbf{Y}_2,\mathbf{Y}_3)\\
&= \frac{1}{2}\log(2\pi e)^n\left|\mathbf{\Sigma}_{Y_1|Y_3}\right|-\frac{1}{2}\log(2\pi e)^n\left|\mathbf{\Sigma}_{Y_1|Y_2}\right|\\
&= \frac{1}{2}\log(2\pi e)^n\left|\mathbf{\Sigma}_{Y_1}-\mathbf{\Sigma}_{Y_1}\mathbf{\Sigma}_{Y_3}^{-1}\mathbf{\Sigma}_{Y_1}\right|-\frac{1}{2}\log(2\pi e)^n\left|\mathbf{\Sigma}_{Y_1}-\mathbf{\Sigma}_{Y_1}\mathbf{\Sigma}_{Y_2}^{-1}\mathbf{\Sigma}_{Y_1}\right|\\
&= \frac{1}{2}\log(2\pi e)^n\left|\mathbf{\Sigma}_{X^*}-\mathbf{\Sigma}_{X^*}\left(\mathbf{\Sigma}_{X^*}+\mathbf{\Sigma}_{\tilde{W}}+\mathbf{\Sigma}_{\hat{W}}\right)^{-1}\mathbf{\Sigma}_{X^*}\right|\\
&\quad -\frac{1}{2}\log(2\pi e)^n\left|\mathbf{\Sigma}_{X^*}-\mathbf{\Sigma}_{X^*}\left(\mathbf{\Sigma}_{X^*}+\mathbf{\Sigma}_{\tilde{W}}\right)^{-1}\mathbf{\Sigma}_{X^*}\right|\\
&= \frac{1}{2}\log(2\pi e)^n\left|\mathbf{\Sigma}_{X^*}\right|\left|I-\left(\mathbf{\Sigma}_{X^*}+\mathbf{\Sigma}_{\tilde{W}}+\mathbf{\Sigma}_{\hat{W}}\right)^{-1}\mathbf{\Sigma}_{X^*}\right|\\
&\quad -\frac{1}{2}\log(2\pi e)^n\left|\mathbf{\Sigma}_{X^*}\right|\left|I-\left(\mathbf{\Sigma}_{X^*}+\mathbf{\Sigma}_{\tilde{W}}\right)^{-1}\mathbf{\Sigma}_{X^*}\right|\\
&= \frac{1}{2}\log(2\pi e)^n\left|I-\left(\mathbf{\Sigma}_{X^*}+\mathbf{\Sigma}_{\tilde{W}}+\mathbf{\Sigma}_{\hat{W}}\right)^{-1}\mathbf{\Sigma}_{X^*}\right|\\
&\quad -\frac{1}{2}\log(2\pi e)^n\left|I-\left(\mathbf{\Sigma}_{X^*}+\mathbf{\Sigma}_{\tilde{W}}\right)^{-1}\mathbf{\Sigma}_{X^*}\right|\\
&= \frac{1}{2}\log(2\pi e)^n\left|I-\left(\mathbf{\Sigma}_{X^*}+\mathbf{\Sigma}_{W}\right)^{-1}\mathbf{\Sigma}_{X^*}\right|\\
&\quad -\frac{1}{2}\log(2\pi e)^n\left|I-\left(\mathbf{\Sigma}_{X^*}+\mathbf{\Sigma}_{\tilde{W}}\right)^{-1}\mathbf{\Sigma}_{X^*}\right|\\
&= 0. \qquad (202)
\end{aligned}$$

If $\left(\mathbf{\Sigma}_{X^*}+\mathbf{\Sigma}_{W}\right)^{-1}\mathbf{\Sigma}_{X^*}=\left(\mathbf{\Sigma}_{X^*}+\mathbf{\Sigma}_{\tilde{W}}\right)^{-1}\mathbf{\Sigma}_{X^*}$, the equality in (202) is satisfied, the equality condition in Lemma 1 holds, and therefore, the equality in (200) is proved. The validity of $\left(\mathbf{\Sigma}_{X^*}+\mathbf{\Sigma}_{W}\right)^{-1}\mathbf{\Sigma}_{X^*}=\left(\mathbf{\Sigma}_{X^*}+\mathbf{\Sigma}_{\tilde{W}}\right)^{-1}\mathbf{\Sigma}_{X^*}$ is proved by Lemma 8 in [27].

Therefore, $I(\mathbf{Y}_1;\mathbf{Y}_2|\mathbf{Y}_3)=0$, and, from the equations in (197), (199), and (200), we obtain the following extremal entropy inequality;

$$\begin{aligned}
&\mu h(\mathbf{X}+\mathbf{V}_G)-h(\mathbf{X}+\mathbf{W}_G)\\
\geq\ &\mu h(\mathbf{X}+\mathbf{V}_G)-h(\mathbf{X}+\tilde{\mathbf{W}}_G)-h(\mathbf{W}_G)+h(\tilde{\mathbf{W}}_G)\\
=\ &\mu h(\mathbf{X}+\tilde{\mathbf{W}}_G+\hat{\mathbf{V}}_G)-h(\mathbf{X}+\tilde{\mathbf{W}}_G)-h(\tilde{\mathbf{W}}_G+\hat{\mathbf{W}}_G)+h(\tilde{\mathbf{W}}_G)\\
\geq\ &\mu h(\mathbf{X}_G^*+\tilde{\mathbf{W}}_G+\hat{\mathbf{V}}_G)-h(\mathbf{X}_G^*+\tilde{\mathbf{W}}_G)-h(\tilde{\mathbf{W}}_G+\hat{\mathbf{W}}_G)+h(\tilde{\mathbf{W}}_G)\\
=\ &\mu h(\mathbf{X}_G^*+\tilde{\mathbf{W}}_G+\hat{\mathbf{V}}_G)-h(\mathbf{X}_G^*+\tilde{\mathbf{W}}_G)-h(\tilde{\mathbf{W}}_G+\hat{\mathbf{W}}_G)+h(\tilde{\mathbf{W}}_G)\\
=\ &\mu h(\mathbf{X}_G^*+\mathbf{V}_G)-h(\mathbf{X}_G^*+\mathbf{W}_G),
\end{aligned}$$

and the proof is completed.

■

REFERENCES

[1] T. M. Cover and J. A. Thomas, *Elements of Information Theory (2nd edition)*, New York: Wiley, 2006.

[2] T. Liu and P. Viswanath, “An Extremal Inequality Motivated by Multiterminal Information-Theoretic Problems,” *IEEE Trans. Inf. Theory*, vol. 53, no. 5, pp. 1839 - 1851, May 2007.

[3] I. M. Gelfand and S. V. Fomin, *Calculus of Variations*, New York: Dover, 1991.

[4] G. Aubert and P. Kornprobst, *Mathematical Problems in Image Processing. Partial Differential Equations and the Calculus of Variations*. Applied Mathematical Sciences vol. 147. Springer Verlag. New York, 2006.

[5] G. Scutari, D. Palomar, F. Facchinei, and J.-S. Pang, Convex Optimization, Game Theory, and Variational Inequality Theory, *IEEE Signal Processing Magazine*, vol. 27, no. 3, pp. 35-49, May 2010.

[6] G. Scutari, D. Palomar, J.-S. Pang, and F. Facchinei, Flexible Design for Cognitive Wireless Systems: From Game Theory to Variational Inequality Theory, *IEEE Signal Processing Magazine*, vol. 26, no. 5, pp. 107-123, Sept. 2009.

[7] J.-S. Pang, G. Scutari, D. P. Palomar, and F. Facchinei, Design of Cognitive Radio Systems Under Temperature-Interference Constraints: A Variational Inequality Approach, *IEEE Trans. on Signal Processing,* vol. 58, no. 6, pp. 3251-3271, June 2010.

[8] M. Debbah and R. Muller, “MIMO Channel Modelling and the Principle of Maximum Entropy, *IEEE Transactions on Information Theory*, vol. 51, no. 5, May 2005.

[9] M. Guillaud, M. Debbah and A. Moustakas, “A Characterization of Maximum Entropy Spatially Correlated Wireless Channel Models, *Information Theory Workshop 2008*, Porto, Portugal.

[10] M. Guillaud, M. Debbah and A. Moustakas, “Modelling the Multiple-Antenna Wireless Channel Using Maximum Entropy Methods, *27th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering,* New York, USA.

[11] D. F. Delong, Jr., and E. M. Hofstetter, “On the Design of Optimum Radar Waveforms for Clutter Rejection, *IEEE Trans. Info. Theory,* Vol. 13, 1967, pp. 454-463.

[12] L. J. Spafford, “Optimum Radar Signal Processing in Clutter*, IEEE Trans. Info. Theory,* Vol. 14, 1968, pp. 734-743.

[13] S. Kay and J. H. Thanos, “Optimal transmit signal design for active sonar/radar,” *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP),* Orlando, Florida, 2002, vol. II, pp. 1513-1516.

[14] E. T. Jaynes, “On the Rationale of Maximum Entropy Methods,” *Proc. of IEEE,* vol. 70, pp. 939-952, 1982.

[15] I. M. Gelfand and S. V. Fomin, *Calculus of Variations*, New York: Dover, 1991.

[16] J. Gregory, *Constrained Optimization in the Calculus of Variations and Optimal Control Theory*, New York: Van Nostrand Reinhold, 1992.

[17] H. Sagan, *Introduction to the Calculus of Variations*, New York: Dover, 1992.

[18] J. Bercher and C. Vignat, “On minimum Fisher information distributions with restricted support and fixed variance,” *Inform. Sci.*, vol. 179, no. 22, pp. 3832-3842, 2009

[19] H. Weingarten, Y. Steinberg, and S. Shamai, “The Capacity Region of the Gaussian Mutiple-Input Multiple-Output Broadcast Channel,” *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 3936 - 3964, Sep 2006.

[20] S. N. Diggavi and T. M. Cover, “The worst additive noise under a covariance constraint,” *IEEE Trans. Inf. Theory*, vol. 47, no. 7, pp. 3072 - 3081, Nov. 2001.

[21] O. Rioul, “Information Theoretic Proofs of Entropy Power Inequalities,” *IEEE Trans. Inf. Theory*, vol. 57, no. 1, pp. 33 - 55, Jan 2011.

[22] A. Dembo, T. M. Cover, and J. A. Thomas, “Information theoretic inequalities,” *IEEE Trans. Inf. Theory*, vol. 37, no. 6, pp. 1501 - 1518, Nov. 1991.

[23] S. Ihara, “On the capacity of channels with additive non-Gaussian noise,” *Inform. Contr.*, vol. 37, no. 1, pp. 34-39, Apr. 1978.

[24] P. P. Bergmans, “A Simple Converse for Broadcast Channels with Additive White Gaussian Noise,” *IEEE Trans. Inf. Theory*, vol. 20, no. 2, pp. 279 - 280, Mar 1974.

[25] T. Liu and S. Shamai (Shitz), “A Note on the Secrecy Capacity of the Multiple-Antenna Wiretap Channel,” *IEEE Trans. Inf. Theory*, vol. 55, no. 6, pp. 2547 - 2553, Jun 2009.

[26] R. Liu, T. Liu, H. Poor, and S. Shamai, A vector generalization of Costas entropy-power inequality with applications, *IEEE Trans. on Information Theory*, vol. 56, no. 4, pp. 18651879, Apr. 2010.

[27] S. Park, E. Serpedin, and K. Qaraqe “An Alternative Proof of an Extremal Inequality,” arXiv:1201.6681.

[28] C. E. Shannon, “A Mathematical Theory of Communication,” *Bell System Tech. J.*, vol. 27, pp. 623-656, Oct. 1948.

[29] S. Verdu and D. Guo, “A simple proof of the entropy power inequality,” *IEEE Trans. Inform. Theory,* vol. 52, pp. 2165-2166, May 2006.

[30] Y. Oohama, “The rate-distortion function for the quadratic Gaussian CEO problem,” *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 1057 - 1070, May 1998.

[31] P. Stoica and P. Babu, “The Gaussian Data Assumption Leads to the Largest Cramér-Rao Bound,” *IEEE Signal Process. Mag.*, vol. 28, no. 3, pp. 132-133, May 2011.

[32] P. Stoica and O. Besson, “Training Sequence Design for Frequency Offset and Frequency-Selective Channel Estimation,” *IEEE Trans. Commun.*, vol. 51, no. 11, pp. 1910-1917, Nov. 2003.

[33] K. Marton, “A Coding Theorem for the Discrete Memoryless Broadcast Channel,” *IEEE Trans. Inf. Theory*, vol. 25, no. 3, pp. 306 311, May 1979.

[34] M. H. M. Costa, A new entropy power inequality, *IEEE Trans. on Information Theory,* vol. 31, no. 6, pp. 751760, 1985.

[35] M. Payaro, M. Gregori, and D. Palomar, “Yet Another Power Entropy Inequality with an Application,” *2011 International Conference on Wireless Communications and Signal Processing (WCSP),* 2011, pp. 1-5.

[36] M. Payaro and D. Palomar, “A Multivariate Generalization of Costa’ s Entropy Power Inequality,” *IEEE International Symposium in Information Theory 2008 (ISIT 2008)*, pp. 1088 - 1092.